Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
NEAL E. MILLER, *Yale University*
KENNETH W. SPENCE, *State University of Iowa*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Assistant Managing Editor
Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS, SARAH WOMACK

CONTENTS

Probabilistic Discrimination Learning: W. K. ESTES, C. J. BURKE, R. C. ATKINSON, AND J. P. FRANKMANN 233
Statistical Learning Theory Applied to an Instrumental Avoidance Situation: A. L. BRODY 240
The Length-Difficulty Relationship in Compound Trial-and-Error Learning: C. E. NOBLE 246
Effect of Pattern Variation Upon Verbal Maze Learning: R. F. THOMPSON, J. F. VOSS, AND W. J. BROGDEN 253
Incidental Learning at Five Stages of Intentional Learning: H. P. BAHRICK 259
Effects of Force and Amplitude Cues on Learning and Performance in a Complex Tracking Task: G. E. BRIGGS, P. M. FITTS, AND H. P. BAHRICK 262
Effect of Variation in Associative Frequency of Stimulus and Response Members on Paired-Associate Learning: G. MANDLER AND E. H. CAMPBELL 269
Studies of Distributed Practice: XVII. Interlist Interference and the Retention of Paired Consonant Syllables: B. J. UNDERWOOD AND J. RICHARDSON 274
Secondary Extinction of Lever-Pressing Behavior in the Albino Rat: W. W. ROZEBOOM .. 280
Error Reinforcement in a Modified Serial Perceptual-Motor Task: M. H. MARX AND R. A. GOLDBECK 288
Conditioning of Muscle Action Potential Responses Resulting From Passive Hand Movement: D. G. DOEHRING 292
Problem Solving in Multiple-Goal Situations: S. B. ANDERSON 297

American Psychological Association
Vol. 54 No. 4 October 1957

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal, experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research effort is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. All manuscripts should be submitted in two complete copies; only the original copy will be returned if the article is not accepted for publication. Figures may be original line drawings or glossy photographic prints, but should not exceed 8½ by 11 in. Care should be taken to insure legibility of labels and critical features of drawings when reduced to a single-column or (exceptionally) a double-column figure in the JOURNAL.

To save printing costs supplementary material to articles in this JOURNAL are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the *Publication Manual of the American Psychological Association*. In particular, the organization of the manuscript should be: text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 54, NO. 4 OCTOBER, 1957

PROBABILISTIC DISCRIMINATION LEARNING[1]

W. K. ESTES, C. J. BURKE, R. C. ATKINSON,[2] AND J. P. FRANKMANN

Indiana University

Traditionally, discrimination experiments have been limited to cases in which two situations to be discriminated differ with respect to some or all of their component stimuli or stimulus attributes (henceforth referred to as *cues*). Analyses of discrimination learning in terms of statistical learning theory (**1, 4**) have led to the study of a more general case which may be termed probabilistic discrimination learning. According to the theory, there is a basis for discrimination learning whenever some of the cues in a situation are differentially correlated with reinforcing events. Classical discrimination learning arises if some of the correlations are unity and others zero; probabilistic discrimination learning arises if the correlations have intermediate values. Otherwise stated, the condition for probabilistic discrimination learning is that differential reinforcement be given relative to two situations which include the same cues and differ only with respect to the sampling probabilities of the cues.

To illustrate these distinctions, consider a standard two-choice discrimination apparatus (e.g., a Grice box or a Lashley stand) with two discriminative cues available, say a light and a tone. To set up a classical discrimination problem, one might present either the light or the tone on each trial and reinforce responses to the left door when the tone is present and responses to the right door when the light is present. To set up a probabilistic discrimination problem, one might assign to the tone and light the probabilities .8 and .4, respectively, of occurring on trials when the left response is reinforced and the probabilities .4 and .8, respectively, of occurring on trials when the right response is reinforced. Now, either the tone or the light or both may occur on any trial, yet there is a basis for better than chance discrimination. According to the theory (**1, 4**), learning relative to the individual cues should proceed exactly as in simple experiments on acquisition with random reinforcement. Thus the asymptotic probability of a given response to any cue taken alone should approximate the conditional probability of reinforcement in the presence of the cue. Using this principle,

[1] This research was supported in part by a grant to W. K. Estes and C. J. Burke from the National Science Foundation.

[2] Now at U.C.L.A.

together with the compounding laws of the statistical model (**1, 3, 7**), it should be possible in advance of an experiment to predict the asymptotic frequency of correct responding and the asymptotic probability of any given response in the presence of any specified combination of cues. For the illustrative example given above, it would be predicted, employing formulae given in a later section, that asymptotically the relative frequency of correct responses would be .56, the probability of a left response to the tone alone would be .67, the probability of a left response to the tone and light together would be .5, and so on (these computations assume equal frequencies of reinforcement on the left and right).

The experiment to be described was conducted to determine whether a probabilistic discrimination would develop under conditions prescribed by the statistical theory and to test quantitative predictions concerning asymptotic response probabilities. A modified verbal conditioning situation offered several advantages for these purposes. Previous studies (**2, 5, 6**) had established that simple acquisition functions for verbal conditioning are described by the statistical model, and the desired stimulus manipulations could be introduced without changing other aspects of the experimental situation.

Method

Apparatus.—The apparatus has been described in detail elsewhere (**5**). The experimental room was dark, except for light coming from the apparatus, and contained four booths and a signal panel. On the panel were mounted 12 12-v., 25-amp. light bulbs, evenly spaced around a circle of 9-in. radius. The booths were placed so that all *S*s would have a clear view of the signal panel at a distance of approximately 7 ft. Within each booth was a pair of telegraph keys, each key directly beneath a reinforcing light. The presentations of signal patterns and reinforcing stimuli were controlled by an automatic programming device and the responses of *S*s were recorded automatically.

Subjects.—The *S*s were 32 non-psychology students obtained from language courses in the Indiana University summer school of 1954 and were paid for their services. They were scheduled in subgroups of four for the experimental sessions.

Design.—For each *S*, one of the reinforcing lights was designated E_1 and the other E_2, the side of the E_1 light being counterbalanced within each subgroup of four *S*s. During discrimination training, either E_1 or E_2 appeared following the signal on each trial; the two types of trials will be designated T_1 and T_2, respectively. On test trials neither of the reinforcing lights appeared.

The stimulating situations obtaining on training trials of types T_1 and T_2 were arranged as follows. First each of the 12 signal lights on the panel was assigned a number between 1 and 12 by a random procedure. Then for T_1 trials the lights in order from 1 to 12 were assigned the probabilities 5/78, 10/78, . . . , 60/78 of being sampled on any one trial; for T_2 trials the lights, in the same order, had sampling probabilities of 60/78, 55/78, . . . , 5/78. In other words, a signal light which had been assigned the ordinal number i appeared on T_1 trials with probability $5i/78$ and on T_2 trials with probability $5(13 - i)/78$, as illustrated in the upper panel of Fig. 1.

During discrimination training, T_1 and T_2 trials occurred in random sequence with the restriction that for one treatment group of 16 *S*s the two types occurred equally often within each period and for the other treatment group of 16 *S*s the proportions were .25 T_1 to .75 T_2. A new random assignment of numbers to signal lights was made for each subgroup of four *S*s in the .5 treatment group and its matched subgroup in the .25 group. The conditional probability of E_1 in the presence of each individual signal light for each group is shown in the lower panel of Fig. 1.

On Days 1 and 2 all groups received 240 training trials. On Day 3 all groups had first a block of 40 additional training trials, bringing the total to 520, then four test blocks of 36 trials each. Each test block included 18 training trials and 18 test trials given in a random order. Each set of 18 test trials included one test on each of the 12 signal lights singly, one test on each of the successive triads (1, 2, 3; 4, 5, 6; 7, 8, 9; 10, 11, 12), and one test each on the mixed triads (1, 3, 11; and 2, 10, 12).

Procedure.—At the first session *S*s were seated in the booths and read the instructions

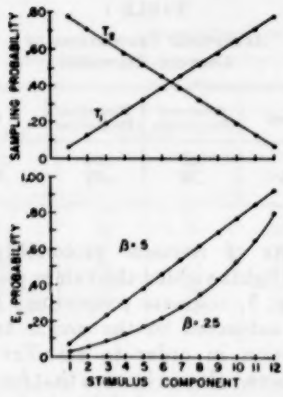

FIG. 1. Upper panel: Sampling probabilities of stimulus components (individual signal lights) on T_1 and T_2 trials (i.e., trials on which reinforcing events E_1 and E_2 occur). Lower panel: Conditional probabilities of E_1 in the presence of individual signal lights for the two treatment groups.

which follow (omitting some repetitious passages):

"I want to see how well you can do in a rather unusual problem situation. At the beginning of each trial some lights will appear on this panel. About a second later, either the left or the right lamp in your compartment will light for a moment. As soon as the signal pattern flashes on the panel, you are to show whether you expect the left or the right lamp in your compartment to light on that trial by pressing the proper key . . . . If you are not sure, guess—your guesses or hunches may turn out to be right, and it is important to have a complete record of your learning.

"Please try not to feel frustrated or irritated if you cannot understand what the experiment is all about. Trying to figure out what the experimenter is up to is not part of your task, and would only tend to spoil your record . . . . The most important thing you can do to ensure the success of the experiment is to indicate as accurately and candidly as possible on each trial which of the compartment lights you expect to appear."

After four practice trials and review of the instructions, the 240 training trials were run off without interruption. Time relations within training trials were approximately: 2-sec. signal duration; 2-sec. delay between signal and reinforcement; 1-sec. duration of reinforcing light; 4-sec. intertrial interval.

On Day 2, the instructions were reviewed and 240 additional training trials were given under the same conditions.

On Day 3, the original instructions were reviewed and the following added:

"Up to now, either the left or the right lamp in your compartment has flashed on each trial to show you which key was the correct response to the pattern shown on the panel. During today's session there will be some trials on which neither of the compartment lights will flash. These test trials will be included to see if you can keep up the same level of accuracy when you are not given information as to whether you are right or wrong on every trial. Aside from this one feature, the situation will not be changed in any way, and you are to continue making your best guess on every trial."

Then the 72 test trials and 112 additional training trials were given in the order described above. On test trials the signal duration and intertrial interval were the same as on training trials.

RESULTS AND DISCUSSION

Statistical analyses.—The key-pressing responses corresponding to *S*'s predictions of E_1 and E_2 will be designated as A_1 and A_2, respectively. In analyses of discrimination learning, the dependent variable will be the mean proportion of correct responses per 40-trial block for a given group; a correct response is defined as an A_1 response on a T_1 trial or an A_2 response on a T_2 trial. In analyses of asymptotic probabilities of response to single lights and to compounds, the dependent variable will be the mean proportion of A_1 responses for a given group on all test trials of a given type. Statistical analyses employed F or t tests whenever possible; when variances could not be stabilized by an arcsin transformation of the scores, nonparametric rank tests (**8**) were used. Statistical significance was defined as the .05 level for a two-tailed t test and the equivalent for other tests.

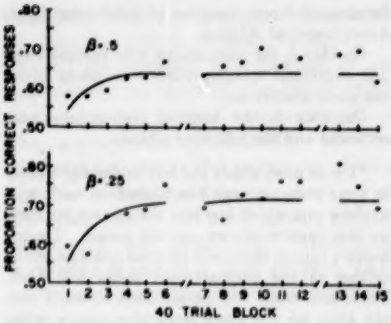

FIG. 2. Theoretical and empirical curves of probabilistic discrimination learning for the two treatment groups. (Note:—the last point on each curve represents only 32 trials.)

Empirical curves of discrimination learning for the two treatment groups are plotted in Fig. 2. Acquisition proceeded at about the same rate in the two groups and was virtually complete by the end of the first period; both curves appear essentially horizontal for the last 400 trials. To estimate the asymptotic probabilities of correct responding, the proportions over the last 40 trials of each period have been averaged. These mean asymptotic proportions are given in Table 1; the group difference is significant by a rank test, and both group means differ significantly from .5 by t tests. Further, both means fall substantially below the values that could be achieved by statistically optimal "strategies."[3]

[3] The optimal strategy would require S to respond to every signal pattern with the response having the higher conditional probability of reinforcement (i.e., the higher probability of being "correct"). The task of calculating exact expected proportions of success for the optimal strategy is beyond our computational resources. However, by analyzing statistically the actual signal-reinforcement sequences which were generated for the present Ss, it is estimated that these maximal proportions are greater than .90 for both of the present groups.

TABLE 1

ASYMPTOTIC PROPORTIONS OF CORRECT RESPONSES

Group	Observed Proportion	Theoretical Proportion	t
.5	.67	.64	.91
.25	.74	.72	.74

Tests of response probability to single lights yielded the values plotted in Fig. 3; response proportions have been subjected to the arcsin transformation in order to stabilize the variances. It can be seen that for each group, response probability is directly related to and, except at the extreme values, virtually equal to the conditional probability of reinforcement. An analysis of variance on the transformed scores yielded significant F's for differences between Lights and between Groups. (The mean square for Ss-by-Lights interaction was used as the error term in the former case, and the between-Ss mean square in the latter.)

Response proportions obtained on the tests with various triads of lights are summarized in Table 2. The

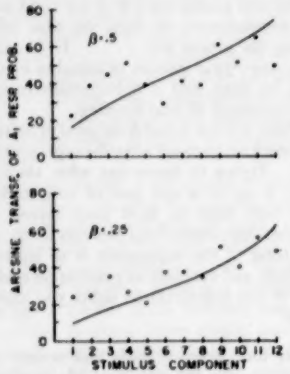

FIG. 3. Predicted and observed probabilities of response to individual signal lights for the two treatment groups.

TABLE 2

PROPORTIONS OF A_1 RESPONSES TO TEST TRIADS

Triads	Proportions				Arcsin Transformations			
	.5 Group		.25 Group		.5 Group		.25 Group	
	Obs.	Th.	Obs.	Th.	Obs.	Th.	Obs.	Th.
1, 2, 3	.22	.15	.03	.06	20	23	4	14
4, 5, 6	.31	.38	.06	.17	29	38	7	24
1, 3, 11	.41	.38	.36	.25	36	38	30	30
7, 8, 9	.59	.62	.42	.35	52	52	38	36
2, 10, 12	.58	.62	.42	.46	51	52	38	43
10, 11, 12	.77	.85	.72	.66	68	67	65	54

triads have been arranged in order of the conditional probabilities of reinforcement, and it can be seen that except for one inversion of an adjacent pair, the observed values fall in the same order for both groups. Group differences are significant, by both t and F tests, for the first two triads, and insignificant for the remainder. Within each group, differences between values for successive triads two or more rows apart are all significant by paired rank tests.

Tests of theoretical predictions.—The theoretical functions required for quantitative predictions of the present data have been derived in detail in another paper (**1**); here the functions will be simply specialized by inserting the particular parameter values called for by the experimental conditions.

In this study the sampling probability of a stimulus element associated with a signal light of ordinal number i is given by

$$\theta_{i\lambda} = \theta(a_\lambda + b_\lambda i), \qquad (1)$$

where λ equals 1 or 2 according to whether the trial in question is of type T_1 or T_2; the parameters a_λ and b_λ are experimenter-determined, but the value of θ must be estimated from data for a particular S or group of Ss. Inserting the appropriate experimental parameters and averaging, it is found that for both types of trials

$$\bar{\theta}_\lambda = 5\theta/12.$$

The theoretical probability of A_1 responses on trials of type T_λ is given by

$$p_{\lambda 1}(n) = \frac{1}{N\bar{\theta}_\lambda} \sum_i \theta_{i\lambda} F_i(n), \qquad (2)$$

where N is the number of cues and $F_i(n)$ is the probability that the ith cue is conditioned to response A_1 after n trials. The latter, in turn, is given by

$$F_i(n) = F_i(\infty) - [F_i(\infty) - F_i(0)] \times (1 - \theta\bar{a} - \theta\bar{b}i)^n. \qquad (3)$$

Under the present experimental conditions, the asymptotic value can be expressed as

$$F_i(\infty) = \frac{\beta i}{(1-\beta)13 + (2\beta - 1)i} \qquad (4)$$

where β represents the relative frequency of T_1 trials and equals .5 and .25 for our two treatment groups. Also, one has, for the present conditions,

$$\bar{a} = (1 - \beta)5/6$$

and

$$\bar{b} = (2\beta - 1)5/78.$$

Assuming the initial probabilities $F_i(0)$ equal to .5, and estimating θ from obtained total response frequencies as in other studies (**2, 5**), Equation 2 has been computed for the two treatment groups. The theoretical curves have been plotted

together with the empirical values in Fig. 2. The values of θ are .05 and .03 for the $\beta = .5$ and $\beta = .25$ groups, respectively. The fits appear reasonably satisfactory, considering that each curve has only one free parameter and that the theoretical asymptotes, depending solely on experimenter-determined parameters, were predictable in advance of the experiment. Moreover, the estimated values of θ are consistent with the values obtained by similar estimation procedures in studies of simple random reinforcement (**2, 5, 6**) and used to predict the data of a previous discrimination experiment (**4**). Whether the apparently systematic deviations of empirical points from the curves on the first few trial blocks of Day 1 and on the terminal blocks of Days 1 and 2 represent a shortcoming of the theory or merely an effect of the particular stimulus sequences used cannot be decided on the basis of the present data. A numerical comparison of predicted and observed asymptotic values is summarized in Table 1.

Predicted asymptotic probabilities of the A_1 response to single lights are given by Equation 4 and illustrated in the lower panel of Fig. 1; it should be noted that these values are also the conditional probabilities of reinforcing event E_1 in the presence of the single lights. The predicted values are compared with corresponding observed values in Fig. 3, both having been subjected to an arcsin transformation. Throughout the middle range, the correspondence of predicted and observed values seems about as good as could be expected in view of the sizes of our experimental groups. At both extremes the observed points deviate in the direction of .5 from the theoretical curves. It may be noted that this is the pattern of deviation to be expected if there are some background cues in the experimental situation which contribute to response probability on all trials. By appropriate analyses one could estimate the weight associated with the background cues and allow for it in the theoretical functions, thus improving the fits of the curves. However, in this study the plan was to deal with theoretical curves predicted in advance from the experimental parameters alone.

For the test of response probability to selected triads of lights, the predicted values are obtained by simply averaging the appropriate F_i values. Consider, for example, the triad of Lights 1, 2, 3. Using Equation 4, one obtains 1/13, 2/13, and 3/13 as the three predicted values of $F_i(\infty)$ for the group which had $\beta = .5$. Averaging these, one obtains $2/13 = .15$, which is the first entry in the column of theoretical proportions for the .5 group in Table 2. The theoretical values thus obtained for all triads tested are compared with the corresponding observed values in Table 2, each pair of values being presented on the left side of the table as proportions and on the right side as angles (arcsin transforms). Of the t tests of differences between theoretical and observed transformed values, those for the first two triads in the .25 group were significant; none of the other t's for the .25 group and none of the t's for the .5 group had P values less than .20. On the whole, the accuracy with which observed probabilities of response to compounds are predicted in this study appears comparable to that achieved by Schoeffler (**7**) for similar tests conducted following classical discrimination learning.

Summary

From statistical learning theory it can be predicted that a new experimental phenomenon, "probabilistic discrimination learning," will be observed if two situations to be discriminated include the same cues and differ only with respect to the sampling probabilities of the cues. Just as partial reinforcement represents a natural generalization of simple acquisition and extinction procedures, probabilistic discrimination learning represents a natural generalization of the conventional discrimination paradigm.

For a test experiment, the two "situations" to be discriminated were constructed by assigning two different patterns of sampling probabilities to a set of signal lights, and a key-pressing response was given differential reinforcement relative to these "situations." Two groups of human *S*s received 592 training trials, under two different reinforcement schedules, and 72 test

trials designed to evaluate asymptotic response probabilities to single cues.

For both groups, curves of correct response proportions approached relatively stable terminal levels intermediate between chance expectations and the maximum values that could be attained by optimal "strategies" under the conditions of the experiment. Rates of probabilistic discrimination learning were very similar to those observed previously for learning under simple partial reinforcement in the same situation. Test trial data confirmed the theoretical prediction that asymptotically the probability of response to any single cue would approach the conditional probability of reinforcement in the presence of the cue.

References

1. Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–146.
2. Burke, C. J., Estes, W. K., & Hellyer, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.
3. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
4. Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.
5. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
6. Neimark, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.
7. Schoeffler, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–329.
8. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.*, 1945, **1**, 80–82.

(Received August 10, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

STATISTICAL LEARNING THEORY APPLIED TO AN INSTRUMENTAL AVOIDANCE SITUATION[1]

ARTHUR L. BRODY[2]

Indiana University

In this study an attempt will be made to extend the statistical learning theory developed by Estes and others (**2, 3, 4, 6, 8**) to an avoidance situation, and to test some hypotheses generated by this extension. The essential condition of avoidance learning is that *S*'s correct response leads to the nonoccurrence of an event.

In the present situation, human *S*s are given a continuous tracking task with one light providing the only illumination for the target. A signal tone is presented every 10 sec. at the start of each trial in a series of discrete trials. Following the onset of the signal, *S* either turns a knob in one direction (Response A_1), turns the same knob in the opposite direction (Response A_2), or fails to turn the knob at all. On a given trial one positive response, A_1 or A_2, is designated the correct avoidance response, and one response, A_1 or A_2, is designated the correct escape response. The correct avoidance response may or may not be the same as the correct escape response. If *S* makes the correct avoidance response within 1.8 sec. of the onset of the signal tone, the light on the target remains on, permitting *S* to continue tracking. In other words, darkness is avoided. The abbreviation, D, will be used for darkness. If *S* fails to make the correct avoidance response within 1.8 sec., D occurs. On any such nonavoidance trial, *S* turns the light back on by making the correct escape response. A correction procedure is followed on nonavoidance trials, and D persists until *S* makes the correct escape response. A trial, then, is either terminated by a correct avoidance response or, when nonavoidance occurs, by a correct escape response.

Both the acquisition of avoidance behavior and the final level of learning are to be studied. The first variable selected which might have an effect on learning is the proportion of trials on which a given response (*a*) leads to avoidance, and (*b*) leads to escape. In previous studies one response has led to avoidance on all trials, and, where escape was permitted (**5, 9**), one response led to escape on all trials. In this situation a response, A_1 or A_2, may lead to avoidance or escape with any specified probability. On any trial, an A_1 is followed by the avoidance of D with probability, π_1, and by the nonavoidance of D with probability, $1 - \pi_1$. Similarly, an A_2 leads to avoidance with probability, π_2, and to nonavoidance with probability, $1 - \pi_2$. An A_1 is followed by escape from D on a proportion, ϕ, of the nonavoidance trials, and since a correction procedure is followed, an A_2 leads to escape on a proportion, $1 - \phi$, of the nonavoidance trials. There will, then, be partial reinforcement at two points. A given response may lead to avoidance less than 100% of the time and, on nonavoidance trials, a given re-

[1] This study is part of a dissertation submitted to the Department of Psychology of Indiana University in partial fulfillment of the requirements for the Ph.D. degree in psychology. The writer is greatly indebted to Professor W. K. Estes, under whose direction the investigation was conducted.

[2] Now at Lehigh University.

sponse may lead to escape less than 100% of the time.

A second variable is the method of terminating the signal tone. Under one condition, called the contingent procedure, the signal is terminated by *S*'s correct response. Under the second, or noncontingent procedure, the signal is terminated by *E* 1.8 sec. after its onset and is not dependent upon *S*'s response.

Estes' statistical learning theory has been developed from experiments in areas other than avoidance learning. This avoidance situation, however, is similar to the experiments on simple human learning (**4, 6, 8**) in all essentials except the relationship between *S*'s response and the nonoccurrence of D. The theory makes the assumption that the stimulus situation present at the start of each trial determines a population of stimulus elements. A proportion, θ, of this population is sampled on any one trial. It follows from the assumption of association by contiguity that these sampled elements become connected to the response which terminates the trial. The probability of a particular response is defined as the proportion of elements in the total population that are connected to that response. When a response leads to avoidance, it is hypothesized that all the sampled elements become connected to that response. This results in an increase in the probability of the response which is followed by avoidance. When a response leads to nonavoidance, it is hypothesized that the status of sampled elements is unchanged, but when this is followed by a response which leads to escape, all the sampled elements become connected to that response. This results in an increase in the probability of the response which is followed by escape. If the hypotheses are correct, the acquisition of avoidance behavior and the asymptotic level of learning will depend upon the relative probabilities with which A_1 and A_2 lead to avoidance and escape.

Method

Design.—The study consisted of a 2×4 factorial design. There were two methods of terminating the signal (the contingent and noncontingent conditions) and four sets of values for π_1, π_2, and ϕ (referred to hereafter as event schedules). Each of the eight groups was run under one of the four event schedules with either the contingent or noncontingent method of signal termination. Six groups, with Event Schedules I, II, and III under contingent and noncontingent conditions, were run first and are designated Exp. 1. The remaining two groups, with Event Schedule IV under contingent and noncontingent conditions, were run later and are designated Exp. 2.

The probability values used for the four event schedules during a first series of 50 trials and a second series of 100 trials are shown in Table 1. All *S*s were given the first series of 50 trials and were required to meet a criterion of at least one avoidance response. Those who failed to do so were discontinued after Trial 50. The *S*s who met the criterion were given the second series of 100 trials. For each *S* three different random and independent sequences of events were drawn, corresponding to the values for π_1, π_2, and ϕ respectively.

TABLE 1

Values of π_1, π_2, and ϕ for the Four Event Schedules

Values	Event Schedules			
	I	II	III	IV
Series 1				
π_1	1.00	1.00	1.00	1.00
π_2	1.00	1.00	1.00	1.00
ϕ	1.00	.66	.34	.30
Series 2				
π_1	.66	.66	.66	.90
π_2	.34	.34	.34	.10
ϕ	1.00	.66	.34	.30

Subjects.—The *S*s were 182 male and female undergraduate volunteers from the beginning course in psychology. They were assigned randomly to the experimental groups. The study was continued until there were 16 *S*s who met the criterion of at least one avoidance response during Series 1 in each of the eight experimental groups.

Apparatus.—The apparatus was a modification of that used by Straughan (**8**). The *S* sat facing a gray panel, 35 in. wide by 24 in. high, which was mounted on a table top. A horizontal slot, approximately at eye level, enabled *S* to view the irregular side-to-side movements of a black spot. The moving spot was actually a segment of an irregular, $\frac{1}{16}$-in.-wide, black line drawn on white paper and moved past the slot by an Esterline-Angus chart drive. The black spot was visible to *S* only when a 7-w. bulb, located above the slot and behind the panel, was lighted. A vertical pointer in front of the black spot could be moved from side to side by means of two rods located below the slot. Above the slot was a knob, kept centered by a small spring and friction device, which could be turned 70° to the left or right before it closed a switch. For a randomly selected one-half of the *S*s, a complete turn of the knob to the left was an A_1 response and a complete turn to the right was an A_2 response. The reverse was the case for the remaining *S*s. An impure tone, approximately 1000 cps, served as the signal.

Procedure.—Instructions to each *S* included the following points: that the pointer was to be kept on the moving spot at all times; that the knob controlled the light on the target and could be turned a considerable distance in both directions; that since the pointer could not be kept on the target if the target could not be seen, the light was to be kept on as much of the time as possible; and finally, that it was possible to keep the light from going off, but if it did go off, to turn it back on as quickly as possible.

Each *S* was then given some practice in turning the light back on as quickly as possible. There were four practice escape trials, during which the signal tone was not presented and *S* could not avoid having the light go off. All *S*s had the same sequence of correct escape responses for the four trials—turning the knob to the left, right, right, and left in that order. After completion of the practice trials, *S* was reminded that during the experiment it was possible to keep the light from going off. Then the experiment began, with a new trial starting every 10 sec. with the presentation of the signal. The tracking task was continuous, and there was no break in going from the first series of trials to the second.

The dependent variable was the occurrence of an avoidance response. The relative frequency of avoidance responses on each trial constituted the measure of response probability. For the first series of 50 trials, since π_1 and π_2 equaled 1.00, either knob-turning response, A_1 or A_2, always led to avoidance if it occurred within 1.8 sec. of the onset of the signal. Both avoidance responses were therefore considered members of the same response class—avoidance responses—and changes in the probability of this class were analyzed. For the second series of 100 trials there was partial reinforcement of each avoidance response. One response class, A_1 avoidance responses, was chosen and changes in the probability of this class were studied for both series of trials. Separate analyses were made of the data from Exp. 1 and the data from Exp. 2.

Results and Discussion

Avoidance responses: Series 1.—In order to obtain 96 *S*s who met the criterion of at least one avoidance response during the first series in Exp. 1, it was necessary to run 135 *S*s. To see whether the 39 *S*s who failed to make an avoidance response were randomly distributed among treatment groups, two separate chi-square tests were run. Significant differences were found among event schedules ($P = .02$) and between methods of signal termination ($P = .01$). It would seem that a greater proportion of *S*s learned to make the avoidance response (*a*) when the escape response was always the same (Event Schedule I), and (*b*) under the contingent condition of signal termination. To obtain 32 *S*s who met the criterion in Exp. 2, 47 *S*s were run. Of the 15 *S*s who failed to learn, 7 had been run under the contingent and 8 under the noncontingent condition.

For *S*s who met the criterion in Exp. 1, a 2×3 analysis of variance for the total numbers of avoidance responses over the 50 trials yielded no significant *F*'s for the two conditions of signal termination, the three event

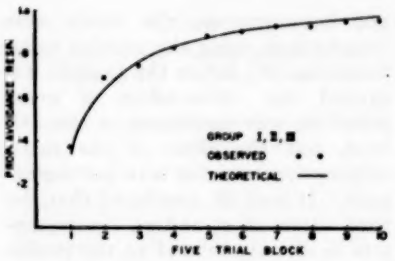

FIG. 1. Proportion of avoidance responses per five-trial block in Series 1 for Exp. 1.

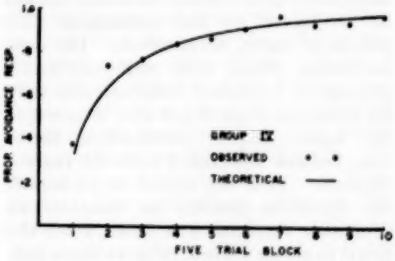

FIG. 2. Proportion of avoidance responses per five-trial block in Series 1 for Exp. 2.

schedules, or the interaction of the two variables. The totals for the two groups in Exp. 2 were within the range of totals for the six groups in Exp. 1.

The learning curves for the eight groups of *Ss* who made at least one avoidance response were negatively accelerated, and ran from an initial value of approximately zero to a final value of approximately unity. Figure 1 shows the learning curve for all *Ss* in Exp. 1 who met the criterion, and Fig. 2 shows the learning curve for all *Ss* in Exp. 2 who met the criterion. These curves were obtained by utilizing methods outlined by Estes and Straughan (**4**). The 96 *Ss* of Exp. 1 were divided into two groups on the basis of learning rate (θ value) and separate theoretical proportions of avoidance responses were calculated for the fast learner group and the slow learner group. These proportions were then averaged, giving one theoretical proportion for each of the 10 blocks of trials. The theoretical curve drawn through these values, together with the observed proportions, is shown in Fig. 1. Since the values for π_1 and π_2 used in Exp. 2 were the same as those used in Exp. 1, there is some basis for assuming that the 32 *Ss* of Exp. 2 should be similar to the 96 *Ss* of Exp. 1 in learning to avoid. To check this, the same θ values estimated from the two groups of 48 *Ss* in Exp. 1 were used as the estimates for fast and slow learner groups of Exp. 2. The satisfactory fit shown in Fig. 2 was obtained.

A_1 avoidance responses: Series 1.— The observed proportions of A_1 avoidance responses in the first series of Exp. 1 for the three event schedules are shown in Fig. 3. In all cases, the proportions are averages for five-trial blocks. To see whether the differences associated with the two independent variables were significant, a 2×3 analysis of variance was computed for total A_1 avoidance responses over the 50 trials in Exp. 1. The effect of event schedules was significant at the .01 level, but the effect of the signal termination variable was

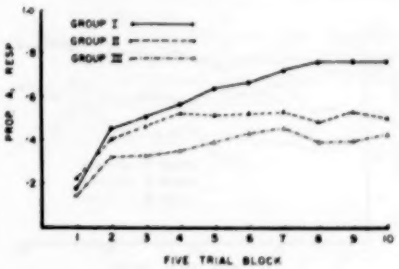

FIG. 3. Proportion of A_1 avoidance responses per five-trial block in Series 1 for Event Schedules I, II, and III.

not significant. It may be concluded that when both A_1 and A_2 lead uniformly to avoidance, i.e., $\pi_1 = \pi_2 = 1.00$, the proportion of A_1 avoidance responses is directly related to the probability, ϕ, that A_1 leads to escape. Additional support for this conclusion is obtained from Exp. 2. The smallest total number of A_1 avoidance responses during Series 1 was found for Event Schedule IV—the group with the smallest ϕ value.

A_1 avoidance responses: Series 2.—The observed proportions of A_1 avoidance responses on Series 2 of Exp. 1 for the three event-schedule groups are shown in Fig. 4. Although the mean curves in Fig. 4 are essentially horizontal, *S*s within each group starting with high or low probabilities showed negatively accelerated learning curves. The probabilities associated with *S*s starting high came down toward the asymptotic value, while the probabilities associated with *S*s starting low rose toward the asymptotic value.

The total number of A_1 avoidance responses on the last 20 trials was selected as the estimate of the asymptote. To determine whether the six groups of Exp. 1 were significantly different with respect to these totals, a 2×3 analysis of variance was run. As the variances of these totals were not homogeneous, the totals were transformed, using the arc-sine transformation (7), before the analysis was carried out. The effect of event schedules was significant at the .001 level, but the effect of the signal termination variable was not significant. It may be concluded that, for fixed values of π_1 and π_2, the asymptote is directly related to the probability, ϕ, that A_1 leads to escape.

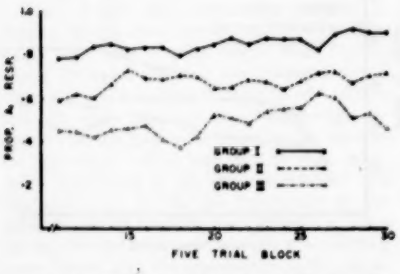

FIG. 4. Proportion of A_1 avoidance responses per five-trial block in Series 2 for Event Schedules I, II, and III.

From a contiguity point of view, the acquisition of avoidance behavior should be facilitated by the contingent procedure of signal termination. On nonavoidance trials with the contingent procedure, a correct response not only led to escape from *D* but also terminated the signal. Signal termination, therefore, became associated with the correct response. If *S* attempted to terminate the signal as quickly as possible, an avoidance response was made, since the signal came on 1.8 sec. prior to the occurrence of D. While the signal termination variable had an effect on the acquisition of avoidance behavior, it had no effect on the proportion of A_1 avoidance responses made by *S*s who learned to avoid. Although the signal was terminated by the correct response with the contingent procedure, signal termination occurred regardless of whether an A_1 or an A_2 was the correct response. Signal termination, therefore, had no influence on the proportion of A_1 avoidance responses.

Consider now the hypotheses that there is an increase in the probability of a response (*a*) when the response leads to avoidance and (*b*) when the response leads to escape. Starting with these hypotheses, the following expression for the mean asymptotic probability value of an A_1 avoidance response $[\bar{p}_1(\infty)]$ can be derived [see (1) for formal derivation].

$$\bar{p}_1(\infty) = \frac{\phi(1 - \pi_2)}{1 - \pi_1 + \pi_1\phi - \pi_2\phi}.$$

One can test the hypotheses by inserting into the above equation the appropriate

TABLE 2

THEORETICAL AND OBSERVED ASYMPTOTIC VALUES FOR THE FOUR EVENT SCHEDULES

Measure	Event Schedule			
	I	II	III	IV
Theoretical	1.00	.79	.50	.79
Observed	.91	.71	.53	.78

values of π_1, π_2, and ϕ for the four event-schedule groups. This gives the predicted asymptotic values which can then be compared with the empirically determined asymptotic values. Table 2 shows these theoretical asymptotes, together with the observed asymptotes based on the total number of A_1 avoidance responses on the last 20 trials. As can be seen, the correspondence is fairly good. The hypotheses selected for avoidance and nonavoidance trials do lead to satisfactory agreement between theoretical predictions and observed results.

SUMMARY

In an avoidance learning situation the variable of partial reinforcement was introduced and the predictive power of Estes' statistical learning theory was tested.

Following the onset of a signal, *S* could make a response which resulted in the avoidance of darkness on a tracked target. If darkness occurred, *S* escaped by making a response which turned the target light back on. Eight groups of 16 human *S*s each were used. The probability that a given response would avoid was varied, as was the probability that a given response would escape on a nonavoidance trial. Termination of the signal was either contingent or not contingent upon a correct response by *S*.

Acquisition of the avoidance response was facilitated by (*a*) uniform, as compared to partial, reinforcement of the escape response on nonavoidance trials and (*b*) the contingent condition of signal termination. The asymptotic level of avoidance learning was directly related to the probabilities with which the given response was reinforced either by avoidance or, on nonavoidance trials, by escape. The forms of acquisition curves and the asymptotic response probabilities predicted from the theory were in good agreement with the data.

REFERENCES

1. BRODY, A. L. Statistical learning theory applied to an instrumental avoidance situation with human subjects. Unpublished doctor's dissertation, Indiana Univer., 1955.
2. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
3. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
4. ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
5. MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54**, No. 5 (Whole No. 247).
6. NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.
7. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State Coll., 1948.
8. STRAUGHAN, J. H. Human escape learning in relation to reinforcement variables and intertrial conditions. *J. exp. Psychol.*, 1956, **52**, 1–8.
9. WICKENS, D. D., & PLATT, C. E. Response termination of the cue stimulus in classical and intrumental conditioning. *J. exp. Psychol.*, 1954, **47**, 183–186.

(Received August 16, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

THE LENGTH–DIFFICULTY RELATIONSHIP IN COMPOUND TRIAL-AND-ERROR LEARNING[1]

CLYDE E. NOBLE

Montana State University

Earlier experiments in this series (**8, 9**) have studied the effects of varying the number of available responses (N_R) in a nonverbal serial multiple-choice task (**11**). Attempting an extension of Hull's (**1**) theory to human trial-and-error learning, the writer (**7, 10**) applied the following rational equation to the acquisition curves:

$$R_p = a(i)r^N, \qquad [1]$$

where R_p = probability of a correct first choice ($1R+$), a = asymptote of R_p (1.00), i = initial probability ($1/N_R$), r = rate parameter related to N_R, and N = number of trials. The average percentage of variance in R_p accounted for was 99.1%.

The present investigation is a further test of the generality of Equation 1. In preceding studies the number of choice points (N_P) was held constant while response availability (N_R) varied. In this experiment N_P and N_R were covaried, as in serial verbal learning. The hypothesis was that the parameters i and r should be sensitive to manipulations of N_P as well as N_R. Related to this was the problem of the regression of difficulty upon length of task. Since the mathematical form of the length-difficulty relationship is contingent upon scale factors, several types of difficulty measure will be analyzed.

[1] This report is based on work done under ARDC Project No. 7707, Task No. 77133, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

S/Sgt. R. B. Card and A 1/C J. B. Downs collected the data in 1953. Statistical analyses were supported by the L.S.U. Council on Research with the assistance of J. H. Bryant, F. J. Farese, D. A. McNeely, M. W. Pryer, and F. E. Stockwell. A report was made to the American Association for the Advancement of Science, December, 1955.

METHOD

Apparatus.—The learning device, which is described in detail elsewhere (**10, 11**), is a semi-automatic 19-key pushbutton apparatus called the Selective Mathometer. The time constants were the same as those used previously (**8, 9**). Overt errors and omissions were scored as $R-$'s, and late choices were recorded but not scored. The stimuli were 14 distinctive electrical circuit symbols chosen for their unfamiliarity to *S*s. These items were prepared as 2 × 2-in. slides to constitute a trial sequence. Stimulus presentation followed an invariant, paced sequence for any given *S*, but the correct responses were required in 12 random permutations for each experimental group.

Four treatments were given, in which the ratios N_R/N_P were 4/4, 6/6, 10/10, and 14/14. The same keys were exposed as in corresponding conditions of the two experiments on response availability (**8, 9**). Considering *S*'s keyboard as being numbered from left to right, the following keys were unmasked for each group:

N_P	Available Keys
4	4,8,12,16
6	4,6,8,12,14,16
10	2,4,6,8,9,11,12,14,16,18
14	1,3,4,5,7,8,9,11,12,13,15,16,17,19

In each condition *S*'s problem was to discover from among all possible solutions the correct sequence of 4 (6, 10, or 14) reactions corresponding to the 4 (6, 10, or 14) stimuli. As before, *S* was instructed to use the forefinger of his preferred hand in making selections and to return to the fixation plate after each response. Repeated choices were allowed during the stimulus period (modified-correction procedure), but *S* was informed that responses made during the

interstimulus or intertrial periods would not be scored. Mastery was defined as five successive trials without error, a more stringent criterion than was employed in earlier research. The reason for increasing the criterion was that approximately 30% of *Ss* who attain two successive perfect trials subsequently make one or more errors. With a criterion of five trials, however, the percentage of backsliders is reduced by half.

Procedure.—A simple randomized design was employed in which the interaction of task length and permutation effects was confounded with individual differences, exactly as in the first two experiments. Each of the four N_P treatments was administered under 12 random response permutations (omitting left-right and right-left sequences), with four *Ss* appearing in each cell. To minimize intraserial effects, the stimulus sequence was changed at random from one group of *Ss* to the next.

The *Ss* consisted of 192 basic airmen at Lackland Air Force Base, assigned randomly to the four experimental treatments within each permutation. They were tested in subgroups of four, and none had prior experience with the Mathometer. For this investigation, four copies of the response panel and recording devices were constructed. The *Ss* were arranged in a semicircle but visually isolated from each other by plywood walls. The problem stimuli were projected on a standard opaque screen to the front, with the equipment located to the rear. Recorder, counter, and reward circuits were, of course, separate for each *S*. When *S* attained mastery he was unobtrusively dismissed from the laboratory by means of a yellow signal lamp on his response panel. Two *E*'s were employed, one to operate the control unit and the other to monitor the recording devices. Following the experiment, each *S* was isolated from the remaining pool of *Ss* in order to prevent discussion of the procedure.

Results and Discussion

The acquisition curves are presented in Fig. 1, where the probabilities (R_p) of correct initial choices ($1R+$) are plotted as functions of trials (N) for the four values of N_P. Consistent with previous work on response availability (**8, 9**), the asymmetrical sigmoid curves are well represented by Equation 1. The best-fitting equations, obtained by minimizing the sums of squares of the residuals to determine the parameter r, are as

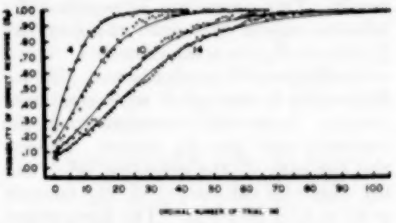

Fig. 1. Probability of initial correct choice (R_p) as a function of practice (N) with number of choice points (N_P) as the parameter. The origins are rational.

follows:

$$N_P = 4: \quad R_p = 1.00(.250)^{.838^N}, \quad [2]$$

$$N_P = 6: \quad R_p = 1.00(.167)^{.904^N}, \quad [3]$$

$$N_P = 10: \quad R_p = 1.00(.100)^{.937^N}, \quad [4]$$

$$N_P = 14: \quad R_p = 1.00(.071)^{.948^N}. \quad [5]$$

The computational procedure begins by taking the second derivative of Equation 1. This gives

$$\frac{d^2R_p}{dN^2} = a(i)^{r^N} r^N (\log_e i)(\log_e r)^2 \times [1 + r^N(\log_e i)]. \quad [6]$$

It can be shown (**2**, p. 130) that the point of inflection (N_i) on the practice dimension is found by setting Equation 6 equal to zero. Solving for N_i we have

$$N_i = \frac{-\log_e(-\log_e i)}{\log_e r} \quad [7]$$

from which it follows that

$$\log_e r = \frac{-\log_e(-\log_e i)}{N_i}. \quad [8]$$

A preliminary estimate of N_i is first obtained by inspection of the acquisition curve. From the ordinate at $1/e = .368$, a perpendicular is dropped from the empirical R_p function to the abscissa. This value of N_i is then substituted in Equation 8, from which an approximation to the rate parameter is obtained. By successive solutions of Equation 1, paired values of observed (R_p) and expected ($\bar{R}_p$) response probabilities

result. This procedure is repeated for selected values of r until the quantity $\sum (R_p - \bar{R}_p)^2$ is a minimum.

Coefficients of determination (R^2) for Equations 2 through 5 were computed directly from the variances of the residuals and the R_p values. Having the variance of the residuals (σ^2_w) and the observed variance (σ^2_0), the formula is $R^2 = 1 - (\sigma^2_w/\sigma^2_0)$. The percentages of determined variance ($100R^2$) are 97.2%, 98.9%, 99.6%, and 99.7% in the order listed. Less distortion results from fitting the curves directly to the raw data than by rectifying Equation 1 via the conventional linear reduction process (cf. **2**), and the present method is also more rapid. As predicted, the i and r parameters jointly account for variations in N_P. There is a related effect on the inflection points; employing Equation 7, zero accelerations occur at 1.81, 5.79, 12.82, and 17.75 trials for Groups 4, 6, 10, and 14, respectively. This positive relationship between N_i and N_P can also be seen by inspection of Fig. 1.

There are some irregularities in the data. The Trial 1 score for Group 4, for instance, shows that these *S*s are performing below chance initially. The empirical R_p value for Group 4 on Trial 1 is .141, whereas the theoretical value is .314 reading from Equation 2. By a z test for proportions the difference of .173 is found to be 2.58 times the $\sigma_{\bar{z}}$ of .067, which is significant at the 1% point.

Using a χ^2 test of goodness of fit (two-category case), the hypothesis of a random distribution of correct responses at the first choice point may be evaluated. The logic of this test (**3**) is that the population of initial choices in Group 4 is distributed according to an asymmetrical binomial: $R_p = .25$, $R_q = .75$. The χ^2 for $df = 1$ is 5.44, which is significant at the 2% point and hence a departure from the hypothesis of chance responding. When the same test is applied to Group 6, the χ^2 is 2.40, which is not significant. Inspection of Fig. 1 indicates that Groups 10 and 14 also respond about as would be expected from their appropriate binomial hypotheses. Thus the likelihood of bias seems to be greater for shorter tasks than for longer tasks. It is an early effect which disappears by Trial 3.

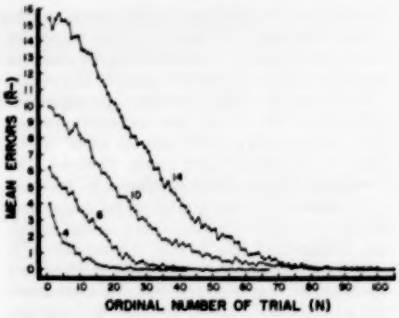

FIG. 2. Mean total errors ($\bar{R}-$) as a function of practice (N) with number of choice points (N_P) as the parameter.

We turn next to an analysis of error curves. In Fig. 2 mean errors ($\bar{R}-$) are shown as functions of trials (N) for the four values of N_P. Except for Group 14 each curve originates at an appropriate point on the ordinate; *S*s in Group 14 make proportionately more errors (overts as well as omissions) on the initial trials than do the other groups. There is also a suggestion of an early increase in errors followed by a later fall for this group.

To any analysis of error data in sequential learning the question of serial position effects is inevitably linked. Since this phenomenon is most frequently observed during early stages of practice (**5**), Fig. 3 presents the mean errors ($\bar{R}-$) cumulated during Trials 1–5 for each group. There is no evidence of the classical function. This horizontal error distribution is typical of Mathometer data (**8, 9**); rather than a bowed primacy-recency effect, there is fairly equal difficulty at each choice point in the series. To evaluate this behavior more fully, independent mixed factorial analyses of variance (**4**, Type I) were performed upon the data in Fig. 3.

Letting the variable A-classifications represent serial positions and the 12 B-classifications represent permutations,

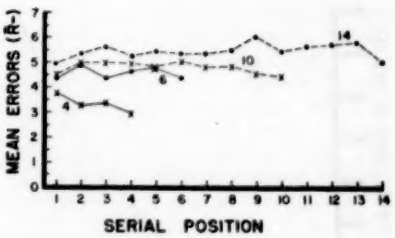

FIG. 3. Mean errors ($\bar{R}-$) during Trials 1–5 as a function of serial position with number of choice points (N_P) as the parameter.

each cell mean was based upon the pooled error scores of four *S*s for the first five trials. Serial position (A) is a significant source of variance for Group 4 ($F = 3.28$; $df = 3/108$) and Group 14 ($F = 2.24$; $df = 13/468$); permutation effects (B) are significant only for Group 6 ($F = 2.07$; $df = 11/36$); and the A × B interactions are significant in all groups ($F \geq 2.11$; $df \geq 33/108$). Primary interest, however, lies in the A-effects. Although there are significant differences among choice points for Groups 4 and 14 in Fig. 3, there is no necessity to test for the typical bowed hypothesis. (An appropriate test might be a comparison of the terminal and center pairs of points.) The only obvious trend is that for Group 4, but since neither of the previous studies (**8, 9**) revealed a gradient for the 4/4 condition, this must be interpreted as experimental error. In other words, the Mathometer response units comprise a chain whose links are approximately equal with respect to variations of N_R and N_P.

Finally, the relationships between "difficulty," variously defined, and length of task must be considered. Table 1 presents five different measures of difficulty as functions of N_P. The scores are defined as follows:

$\dot{N}$ = median total trials to mastery, computed from the trial number on which each *S*'s mastery run is *begun;* i.e., the first of 5 perfect trials.

$\bar{R}-$ = mean initial errors; i.e., the average number of errors cumulated during Trials 1–5 including omissions.

$\dot{R}-$ = median total errors to mastery, recorded over the same trials as $\dot{N}$.

r = rate parameter, calculated by empirical curve-fitting in Equations 1 and 8.

N_i = inflection point, computed from Equation 7 as the trial number on which acceleration is zero.

In the case of each type of score in Table 1, the index of difficulty increases with N_P. None of the measures, however, is appropriate for determining the law governing the amount of work required to learn different amounts of the same material. As McGeoch (**5**) pointed out, trials-to-criterion scores fail to hold temporal duration constant; thus, although $\dot{N}$ is a positive (linear) function of N_P, the units of measurement change systematically from group to group. Initial errors are also unsatisfactory; the amount of practice is comparable but a uniform performance criterion for $\bar{R}-$ is lacking. The $\dot{R}-$ measure meets the foregoing objections because total errors are correlated with total time; unfortunately, however, $\dot{R}-$

TABLE 1

VARIOUS PERFORMANCE MEASURES OF DIFFICULTY AS FUNCTIONS OF THE NUMBER OF CHOICE POINTS (N_P)

N_P	Median Total Trials ($\dot{N}$)	Mean Initial Errors ($\bar{R}-$)	Median Total Errors ($\dot{R}-$)	Rate Parameter (r)	Inflection Point (N_i)
4	10.0	13.5	20.0	.835	1.81
6	22.5	27.6	67.0	.904	5.79
10	40.0	47.9	234.0	.937	12.82
14	60.5	76.6	473.5	.946	17.75

is difficult to rationalize since it is so strongly affected by individual differences. Rate parameters provide a more rational index of difficulty, but r is subject to the same criticism as $\bar{N}$ with respect to the problem of equal units. The N_t concept, despite its abstract properties and the possibility of coordinating it with Equation 1, must also be rejected on similar grounds. The most suitable index, derivable from the trials-to-criterion measure, is the total time required to reach mastery. Here the units are equal, the criterion is uniform, and the dimension has a meaningful origin.

Perhaps the most precise hypothesis about length-difficulty phenomena is one which Thurstone (13) derived from his rational equation (12) for the verbal learning function:

$$T = c/k(N_P)\sqrt{N_P - s}, \qquad [9]$$

where T = total time to criterion, c = E's criterial constant, k = S's learning constant, N_P = number of choice points, and s = S's span of attention.

By making some simplifying assumptions, it can be shown that Equation 9 as applied to the Mathometer is essentially a parabola. First of all, the concept of attention span (s) is not relevant to trial-and-error learning, so we may set $s = 0$. Second, there should be a meaningful T value for a single-link problem, and it would be desirable to have the curve originate at zero. Since only by chance can S make a perfect score in any condition on Trial 1, we may let the time required for a single predetermined choice be equal to the stimulus duration plus the interstimulus and intertrial periods. The constant ratio c/k, therefore, may be replaced by the (log intercept) parameter a specifying the unit time of work and rest. Finally, since $N_P\sqrt{N_P} = N_P^{1.5}$, we may assume more generally that the equation we seek has a slope parameter b whose value is greater than unity, thus producing positive acceleration as in Thurstone's (13) experimental examples.

From the foregoing, it follows that a simple logical adaptation of Equation 9 to the present situation would be

$$\hat{T} = a(N_P)^b, \qquad [10]$$

where $\hat{T}$ = median total time to mastery in seconds, calculated for each N_P condition as the median trials to mastery ($\bar{N}$) multiplied by the trial duration, i.e., $\bar{N}[N_P(2.00 + 1.44) + 6.88]$; a = trial duration in seconds for $N_P = 1$, calculated as the sum of the stimulus, interstimulus, and intertrial periods, i.e., 10.32 sec.; and b = slope parameter > 1.0, inversely related to ability and to distribution of practice.

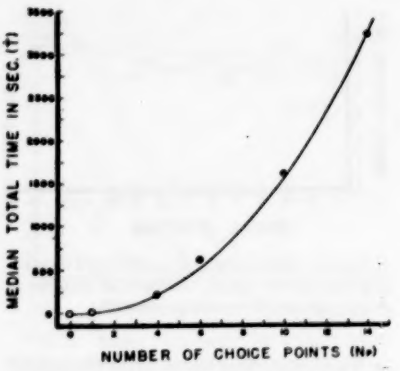

FIG. 4. The length-difficulty relationship in compound trial-and-error learning. The first two points are rational.

By causing the curve in Fig. 4 to pass through the rational points (0,0) and (1, 10.32), then minimizing the sums of squares of the remaining four residuals to determine b, the best-fitting parabolic equation is

$$\hat{T} = 10.32(N_P)^{2.19}. \qquad [11]$$

The per cent of variance in difficulty ascribable to Equation 11 is 99.8%.

The major hypothesis of the present experiment may be considered verified; i.e., that the parameters i and r in Equation 1 both respond to joint variations of task length and response availability. Furthermore, it appears that the basic form of probability acquisition curves in human trial-and-

error learning is invariant over changes in the number of choice points. What is denoted here by "basic form" is the general mathematical equation (exclusive of the specific numerical values of its parameters) which most closely approximates the empirical data; i.e., Equation 1. This hypothesis will be recognized as a logical consequence of the Kjerstad-Robinson law (5, p. 48).

There remains the problem of the regression of "difficulty" upon length of task. The present analysis, in agreement with McGeoch (5) and Thurstone (13), indicates that total time to mastery is the most appropriate index of difficulty, and that a rational equation for this function has the parabolic form $Y = aX^b$, where $b > 1$. Assuming, therefore, that the length-time relationship for compound trial-and-error learning is of the type described by Equation 10, it is interesting to compare the results of a similar investigation of perceptual-motor learning by the writer (6). The apparatus was the Complex Coordination Test. The variable N_P was manipulated in terms of the number of successive visual patterns to be matched, and difficulty was expressed as total time per match. In that experiment 99.9% of the variance was determined by a *linear* equation. Although the differences between trial-and-error learning and perceptual-motor skill are many, both experiments involved the acquisition of an invariant sequence of paired associates. That is, each response in the series was connected to a unique stimulus, and the reinforcing event was independent of the preceding response and of the succeeding stimulus.[2]

As was previously discussed (6), positive acceleration is a characteristic of "competitional" learning problems. whereas "noncompetitional" situations give linear relationships. In the former case, time per unit is contingent upon task length, but in the latter it remains constant. Thus, despite its freedom from classical serial position effects, the Selective Mathometer is a sequential learning device of the "competitional" or habit interference type. Apparently, however, the successive acts in a compound chain may compete with each other for order of appearance without producing differential probabilities of error at the several choice points. The practical importance of this is that one may investigate the dependence of difficulty upon complexity in the absence of intraserial complications.

[2] In mazes and serial verbal learning these factors are confounded, with the result that anticipatory and perseverative errors occur most frequently in the middle of the chain.

Summary

This experiment involved compound trial-and-error learning under four conditions of task length (N_P). The *S*s ($N = 192$) practiced series of either 4, 6, 10, or 14 pushbutton responses to a criterion of five perfect trials.

Response probability curves for the four N_P conditions were skewed and S-shaped. Each curve originated at a hypothetical R_P value consistent with the initial chance solubility (i) of the problem. The inflection points (N_i) were a direct function of N_P, and the acquisition rates (r) varied inversely. The rational equation $R_P = a(i)^{r^N}$ accounted for an average of 98.8% of the variance.

With respect to the regression of difficulty upon length of task, six different types of score were analyzed. The most appropriate index was the median total time to mastery in seconds ($\tilde{T}$). The rational equation $\tilde{T} = a(N_P)^b$ accounted for 99.8% of the variance. Consistent with earlier research, errors were practically independent of serial position but there were significant permutation by position interactions.

The findings were discussed in relation to Thurstone's length-difficulty function and the Kjerstad-Robinson law.

References

1. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
2. Lewis, D. *Quantitative methods in psychology.* Ann Arbor: Edwards, 1948.
3. Lewis, D., & Burke, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949, **46**, 433–489.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.

6. Noble, C. E. Amount set and the length-difficulty function for a self-paced perceptual-motor skill. *J. exp. Psychol.*, 1953, **46,** 435–439.

7. Noble, C. E. The difficulty-complexity relationship in compound trial-and-error learning. *Amer. Psychologist,* 1954, **9,** 443. (Abstract)

8. Noble, C. E. Compound trial-and-error learning as a function of response availability (N_R). *J. exp. Psychol.*, 1955, **49,** 93–96.

9. Noble, C. E. Supplementary report: response availability in human trial-and-error learning. *J. exp. Psychol.*, 1957, **53,** 359–360.

10. Noble, C. E. Human trial-and-error learning. *Psychol. Rep.*, 1957, **3,** 377–398 (Monogr. Suppl. 8).

11. Noble, C. E., & Farese, F. J. An apparatus for research in human selective learning. *J. Psychol.*, 1955, **39,** 475–484.

12. Thurstone, L. L. The learning function. *J. gen. Psychol.*, 1930, **3,** 469–493.

13. Thurstone, L. L. The relation between learning time and length of task. *Psychol. Rev.*, 1930, **37,** 44–53.

(Received August 24, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

EFFECT OF PATTERN VARIATION UPON VERBAL MAZE LEARNING[1]

R. F. THOMPSON, JAMES F. VOSS,[2] AND W. J. BROGDEN

University of Wisconsin

In previous studies on the effect of number of alternate choices per unit of a verbal maze on acquisition and serial position errors, note was made of an apparent effect of the pattern of the series of correct responses (**1, 2**). Even though the maze pattern was generated by use of a table of random numbers, acquisition and frequency of serial position errors appeared to be altered differentially by certain restricted sequences within the total pattern. Since attempts to provide a quantitative scale for pattern characteristics by means of mathematical models have not to date been successful, empirical approaches to systematic variation of pattern characteristics were considered. One such approach involves systematic variation of the location of a doublet (two successive like responses) within the maze pattern. The present paper reports the results of an experiment designed to investigate the hypothesis that systematic variation in the location of a doublet within the maze pattern will affect acquisition and serial position errors.

METHOD

The 16-unit maze with four alternate choices, each occurring four times in the pattern, used by Brogden and Schmidt (1) was altered in the following manner. The identical pattern was used for the control group except that the alternative choices were the numbers 10, 20, 30, and 40. For the five experimental groups, the basic maze pattern was changed to provide a doublet for Maze Units 2 and 3, 5 and 6, 8 and 9, 11 and 12, and 14 and 15. The pattern for each one of the experimental groups and that for the control group was permuted by first adding 10 to each number of the group pattern, second by adding 20, and third by adding 30. The numbers 50, 60, and 70 were reduced to 10, 20, and 30, respectively. The four subpatterns thus obtained provide for the representation of each number once in each of the 16 positions. The doublet in each experimental group pattern is represented by each of the following pairs: 10–10, 20–20, 30–30, and 40–40.

There were 24 *Ss* in each group and six *Ss* of each group learned one of the four subpatterns provided by the permutations of the numbers. Half of the *Ss* in each condition were handled by each of two *Es*, so that *E* was orthogonal to the other experimental conditions. The *Ss* were members of the classes in elementary psychology at the University of Wisconsin who volunteered to serve in an experiment on learning a verbal maze. The *Ss* were assigned to the various conditions by a random procedure.

The following instructions were read by *E* to each *S*. "In this experiment you will be required to learn a single list of numbers. These are the numbers used—10, 20, 30, and 40. Each number is used more than once in the list. When I tell you to begin, start naming these four numbers. When you name the correct number for the first position of the list, I will say 'forward.' Then start naming the numbers again. When you name the second correct number of the list, I will say 'forward.' This procedure will be continued until the end of the list is reached. Please respond as fast as you can. When the first trial is completed, there will be a short rest before the second trial is started. We will continue with similar trials until you are able to make one repetition of the list without error. Are there any questions?"

After answering any questions, *E* said "Begin" and *S* responded by giving one of the four alternate responses. The procedure described in the instructions was followed until *S* attained the criterion of one errorless trial. The *E* recorded every response for each unit on each trial. A stopwatch was used to measure the time of each trial and the one-minute rest

[1] This research was supported in part by a grant from the Graduate Research Committee from funds provided by the Wisconsin Alumni Research Foundation.

[2] Wisconsin State College, Eau Claire.

interval between trials. Other conditions of the experiment were comparable to those used in previous studies (1, 2).

Results

The number of trials, the total errors, and the total time to the criterion trial were tabulated for each *S*. An analysis of variance was performed on the four permutations of the maze pattern for the control group and for each of the five experimental groups for the three measures of acquisition. Since no significant differences between permutations were found for any of the groups, the data from the permutations were pooled for further analysis of the acquisition data. An analysis of variance for each of the three sets of acquisition data was performed and in no case was maze pattern (doublet position) a significant source of variation, nor were there any other significant differences.

Mean group error over all trials to the criterial trial was computed for each group to obtain the data for plotting the filled-circle solid-line serial position error curves presented in Fig. 1. Examination of these curves indicates that those for the experimental groups are of different form from each other and from that of the control group. When comparison is made between each curve of the experimental groups and the curve for the control group at the locus of the doublet in the curve of the experimental group (represented by the two vertical arrows with unbroken connection) the frequency of error at the position of the second item of the doublet is greater than is the frequency of error at the position of the first item of the doublet. This phenomenon is more directly shown by the open-circle broken-line curves plotted for each of the experimental groups. These represent the difference between the serial position error curves of the appropriate experimental group and the control group. Since irregularities of similar nature to those apparently produced by the doublet were noted in each of the curves for the experimental groups, an examination was made of the maze patterns. For nearly all cases of substantial deviation of the nature apparently produced by a doublet, it was found that a given response item was followed by an identical response item one position removed, e.g., 20—10—20 for the Positions 13, 14, and 15 of the maze pattern for Group E1. This sequence form we have called a split-doublet. There are no split-doublets in the pattern of the control group. Their occurrence in the patterns of the experimental groups is a function of the method of altering the pattern of the control group in order to provide for the location of a doublet at different serial positions in the experimental group patterns.

The differences between the five experimental group curves representing the difference of each *S* in the experimental groups from the control group curve were tested by an analysis of variance, the results of which are summarized in Table 1. Serial Position, the interaction of Serial Position and Maze Pattern, and the triple interaction of Serial Position, Maze Pattern, and Maze Pattern Subgroups are the only significant sources of variation.

The significance of the Serial Position source of variation indicates that the points of the over-all curve differ as a function of serial position. The interaction of Serial Position and Maze Pattern indicates that the form of the Serial Position curves varies as a function of maze patterns or groups. The differences between curve form

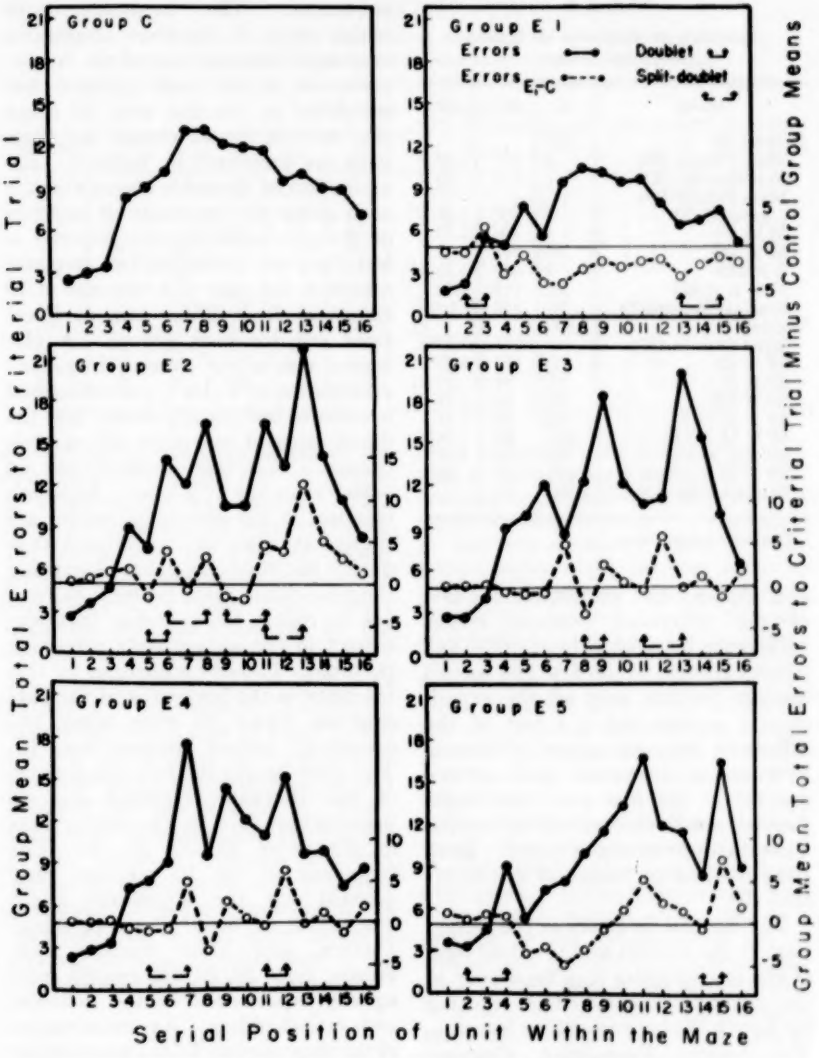

FIG. 1. Serial position error curves as a function of maze pattern.

include those produced by the locus of both the doublet and split-doublet sequences, since these two variables are confounded. The significant triple interaction indicates that the serial position curves for the different maze patterns differ as a function of the maze pattern subgroups.

In order to obtain tests of the effects apparently produced by the doublet

TABLE 1

ANALYSIS OF VARIANCE OF SERIAL POSITION ERRORS

Source	*df*	*MS*	*F*
Between *Ss*	119		
Maze Pattern (M)	4	877	2.45
Experimenter (*E*)	1	118	.33
Maze Pattern Subgroup (MS)	3	680	1.90
M × *E*	4	498	1.39
M × MS	12	267	.75
E × MS	3	609	1.70
M × *E* × MS	12	185	.52
Residual Between *Ss*	80	358	
Within *Ss*	1,800		
Serial Position (SP)	15	273	7.37*
SP × M	60	172	4.65*
SP × *E*	15	34	.92
SP × MS	45	44	1.19
SP × M × *E*	60	42	1.14
SP × M × MS	180	46	1.24*
SP × *E* × MS	45	39	1.05
SP × M × MS × *E*	180	15	.41
Residual Within *Ss*	1,200	37	

* Significant at .05 level.

and split-doublet sequences, a *t* test of the difference between means (difference between experimental and control groups) of the first and second doublet position over all the experimental groups and a *t* test of the difference between means (difference between experimental and control groups) of the first and third split-doublet positions over all the experimental groups were performed. Both tests resulted in values of *P* smaller than .05.

The increase in errors at the second item of the doublet and the third item of the split-doublet may be similar to the negative recency effect reported by Jarvik (**3**) for probability learning. The present four-choice alternate response and maze length represent a marked difference from Jarvik's two-response guessing situation and series length. An analysis of the first trial of the present *Ss* provides a test of whether their guessing behavior contributes to the effect of the doublet sequences. The frequency with which each of the four alternative responses followed each of the correct responses in the maze pattern was tabulated on the first trial for every *S* of each of the six groups and these data are presented in Table 2. Examination of the table shows that for each group the frequency of response of the previously correct response is less than any other kind of response except in the case of a response of 10 (diagonal from upper left to lower right cell for each group). A chi-square test of the over-all frequency distribution of Table 2 was performed in order to test the hypothesis that the distribution of responses differs from chance. This test is significant at better than the 5% level. Although this test is not direct, its results are consistent with the hypothesis that during the discovery phase of learning the four-choice maze pattern, *Ss* tend not to repeat the response that was correct in the immediately preceding position. The one exception to this tendency is the frequency of the first response being 10 even when the previously correct response was 10. This may be due to the ordinal series of the alternate responses and its presentation to *S* in the instructions in the order 10, 20, 30, 40. The frequency of the 10 response may account for the significant triple interaction of Serial Position, Maze Pattern, and Maze Pattern Subgroups, since 10, 20, 30, and 40, were equally represented in the doublets and split-doublets. An examination of the responses on Trial 2 gave similar results, except that there was a decrease in the frequency of 10 as the first response following any alternative. This change may be attributed to effects from the first trial.

The first-order contingent responses to the third element of all three

TABLE 2

FREQUENCY OF RESPONSE ON THE FIRST TRIAL FOR EACH OF THE FOUR ALTERNATIVE RESPONSES FOLLOWING EACH CORRECT RESPONSE OF THE MAZE PATTERN

Prior Correct Response	First Order Contingent Response											
	10	20	30	40	10	20	30	40	10	20	30	40
	Group C				*Group E1*				*Group E2*			
10	46	24	13	7	45	30	8	8	43	23	15	12
20	64	1	18	7	48	7	21	14	42	7	25	15
30	53	18	0	19	44	7	12	26	47	13	5	25
40	72	6	11	1	53	14	13	10	51	19	10	8
	Group E3				*Group E4*				*Group E5*			
10	47	22	12	9	30	32	12	16	32	31	19	8
20	50	1	23	16	43	6	22	19	37	3	24	26
30	54	15	3	18	45	14	4	27	46	19	6	19
40	59	11	17	3	49	13	21	7	55	13	13	9

element sequences were tabulated for the first trial for each *S* of all groups in terms of the first element. This frequency distribution was tested for the split doublet effect by chi square. Although the test was significant at the 5% level, the distribution differs from that obtained previously for the two-element sequences used in testing the doublet effect. There is a high frequency of responses of 10 similar to that noted in the data of Table 2, but there is no decrease in frequency of response along the diagonals as was found in the data of Table 2. Thus, there is no indication that the *S*s tend not to repeat a response that was correct for the maze unit two positions backward. Since the presence of the split-doublets in the maze patterns was not planned, no other tests of the split-doublet effect were attempted.

DISCUSSION

The discovery of the effect of the split-doublet sequence upon serial position errors has revealed the experimental design of the present study to be faulty in terms of its intentions. Instead of providing for an analysis of the effect of maze pattern upon acquisition and serial position error curves as a function of the locus of doublets, the independent variable of maze pattern represents a confounding of the presence and position of doublets and split-doublets. The failure to find any significant effect of maze pattern upon trials, time, or errors as acquisition measures may be due to interactions between the doublet and split-doublet variables. The significant effect of maze pattern upon errors to the criterial trial for the experimental groups minus the errors to the criterial trial for the control group suggests that doublet position, split-doublet position, or both may have an effect upon measures of acquisition.

The significant effect of doublet and split-doublet position upon the form of the serial position error curve is clearly established. Although the over-all form of the serial position error curve is a function of the presence and loci of doublet and split-doublet sequences, the effect in both cases is one of an increase in error. In the case of the doublet sequence, the second item of the doublet produces a frequency of error significantly greater than does the first item. In the case of the split-doublet sequence, the third item of the sequence produces a significantly greater frequency of error than does the first item. A more precise determination of these effects will result, of course, from experiments

in which doublet and split-doublet position are manipulated separately as independent variables.

The analysis of the first order contingent responses indicates that the guessing habits of *S*s and the instructions given *S* about the sequences of guessing may be critical factors in the acquisition of alternate choice maze patterns, particularly during the discovery phase. The interaction of these variables with variables of maze pattern complicate investigations of the effects of the latter variable on acquisition and serial position errors. It may be hypothesized, however, that the interaction of these two kinds of variables may result in facilitation as well as in a decrement of acquisition of verbal mazes.

Summary

This experiment was designed to test the effect of locus of a doublet (two successive like responses) on the acquisition and serial position errors of a 4-choice, 16-unit verbal maze. The maze pattern for the control group had no doublets. Those for the five experimental groups had a single doublet at maze units 2 and 3, 5 and 6, 8 and 9, 11 and 12, and 14 and 15, respectively. Analysis of variance of the acquisition data in terms of trials, time, and errors to the criterial trial resulted in no significant differences. Examination of serial position errors in terms of the difference between the control group and each of the experimental groups revealed an apparent effect of an increase in error for the second item of the doublet. A similar effect was discovered for split-doublets (two identical response items with a different response item between them) in that frequency of error was greater for the second one of the two identical responses. The split-doublet sequences occurred due to alteration of the pattern for the control group to provide for the doublets in the maze patterns for the experimental groups. Analysis of variance of the serial position error data showed significant differences in the form of the curves. Further analysis by *t* tests revealed the doublet and split-doublet effects to be significant. Analysis of first order contingent responses suggests that *S*s tend not to repeat a response that was previously correct. This kind of guessing behavior may account for the doublet effect.

References

1. Brogden, W. J., & Schmidt, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.
2. Brogden, W. J., & Schmidt, R. E. Acquisition of a 24-unit verbal maze as a function of number of alternate choices per unit. *J. exp. Psychol.*, 1954, **48**, 335–338.
3. Jarvik, M. E. Probability learning and a negative recency effect on the serial anticipation of alternate symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

(Received August 27, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

INCIDENTAL LEARNING AT FIVE STAGES OF INTENTIONAL LEARNING

HARRY P. BAHRICK [1]

Ohio Wesleyan University

This study investigates the acquisition of incidental learning at progressive stages of intentional learning. Specifically it is hypothesized that incidental and intentional learning are complementary, so that incidental learning occurs primarily at those stages of intentional learning when motivation is relatively low, or when *S* is uncertain of the task set by *E*.

Several recent studies (**2, 4, 5**) have indicated that incidental learning is subject to many of the same effects observed with intentional learning. The progress of incidental learning in relation to the simultaneous progress of intentional learning has, however, not been investigated. In one earlier study (**1**) it was shown that *S*s given a high incentive for the task set by *E* will acquire less incidental learning than *S*s provided with a lower incentive. These findings were related to cognition type learning theory by the following rationale: The *S* is usually under the simultaneous influence of several goal demands, and only one of these may be directed toward the incentive provided by *E*. If the several goal demands are nearly equal in strength, one may expect frequent changes of set. If one of the goal demands greatly exceeds the others in strength, less variability of *S*'s set is to be expected. Further, it may be assumed that "intentional learning," or learning relevant to the task defined by *E*, occurs whenever the corresponding set is present, and incidental learning whenever *S* responds with another set determined by another goal demand. When the incentive offered by *E* is increased, one would expect a stronger goal demand in relation to it, therefore a more persistent set and less incidental learning. With a smaller incentive a weaker goal demand would be expected, therefore greater variability of set and more incidental learning.

This same rationale would lead to the prediction that the rate of progress of incidental learning is quite different from, and perhaps somewhat complementary to, the progress of intentional learning. More specifically, it has generally been assumed that the strength of an incentive increases as it is approached, and that *S*s show more persistent efforts directed toward mastery of a task as they near the completion of the task. This would suggest that relatively little incidental learning occurs during the later portions of the course of intentional learning, and relatively more during the early portions.

The present experiment was designed to explore empirically the relations between the course of intentional and the course of incidenal learning.

METHOD

Subjects.—One hundred *S*s satisfying a course requirement for elementary psychology were assigned alternately to five groups of 20 *S*s each.

Apparatus.—The apparatus used was a memory drum which exposed serial lists at a 3-sec. rate; a practice list consisting of seven meaningful adjectives, and an experimental list

[1] The author wishes to express his appreciation to Miss Mary Lou Walker and Miss Carolyn Fisher who helped in the collection and analysis of data.

consisting of 16 nonsense syllables. Each of the syllables was enclosed in a different geometric form, such as a circle, square, rectangle, etc. The geometric forms were drawn with India ink on a white background, and each form occupied an area of approximately ¼ sq. in.

Procedure.—All *S*s were first trained on the practice list, with instructions to anticipate each word before it appeared in the drum window, beginning with the second trial. All *S*s were trained on the practice list to the criterion of one perfect trial.

After a rest period of 3 min., all *S*s were instructed to learn the experimental list in the same manner. Instructions were added to spell out each syllable, and no mention was made of the forms in which the syllables were enclosed.

Group I learned to the criterion of four correct responses (25% of the list); Group II learned to a criterion of eight correct responses (50%); Group III to a criterion of 12 correct responses (75%); Group IV to a criterion of 16 correct responses (100%). The *S*s of Group V each learned to a criterion of 16 correct responses plus half the number of trials required to reach that criterion (150%).

Immediately after training was terminated, *S*s of all groups were shown a 16-row by 16-column chart. Each row of the chart showed a particular syllable of the list enclosed in all of the possible 16 geometric forms used in the list. The syllables were arranged along the columns in the same sequence as during learning, and the sequence of forms used within each of the rows was randomized. The *S*s were instructed to point out the geometric form associated with each syllable.

Results

Figure 1 shows incidental learning at five stages of intentional learning. Since there were 16 possible forms we may assume that chance would account for a total of one correct form identification for all of the 16 syllables. On this basis it is apparent that all groups showed significant incidental learning. It further appears that the greatest increments of incidental learning occurred during early (Group I) and late (Group V) stages of practice, with little difference among the intermediate stages of practice.

To test the over-all significance of differences among the groups, the incidental learning data were subjected to an analysis of variance. The results indicate that the over-all group differences are significant ($F = 3.68$, $P < .01$ for df 4 and 95). Individual t tests indicate that Group V (150%) differs significantly ($P < .05$) from all of the other groups, but the remaining groups do not differ significantly from one another.

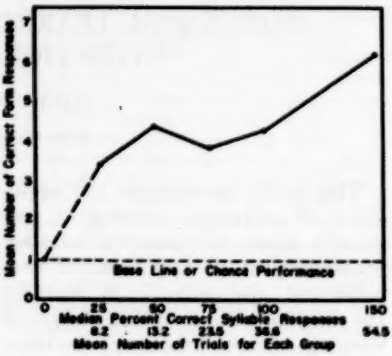

FIG. 1. Incidental learning at five stages of intentional learning.

It is of interest to determine the relation between the degree of intentional learning of a particular syllable and the incidental learning of the geometric form in which that syllable was enclosed. To determine this relation the correct number of responses given to each syllable was summed for *S*s of each group for all trials. Similarly, the total number of correct geometric form responses of each group was determined for each syllable. The correlations between these summed scores for intentional and for incidental learning associated with particular syllables were determined for each group. None of these correlations was significant at the .05 level; they were .28, − .03, − .15, .07, and .30, respectively, for the five groups, suggesting considerable independence of the degree of intentional and incidental learning associated with particular syllables.

Post-experimental report revealed that a total of seven *S*s attempted to memo-

rize the forms despite the fact that the forms were not mentioned in the instructions. These *S*s and their incidental learning scores are distributed as follows: Group I, two *S*s with scores of 6, 8; Group II, one *S* with a score of 10; Group III, three *S*s with scores of 4, 7, and 5; Group IV, none; Group V, one *S* with a score of 12.

Discussion

The results suggest different types of acquisition functions for intentional and incidental learning connected with a particular task. In accordance with prediction it was found that little or no incidental learning occurred between the 50% and the 100% level of intentional learning. At this stage of intentional learning there is reason to expect that the most persistent efforts are directed toward mastery of the task set by *E*. The large amount of incidental learning during the very early trials has been attributed to greater variability of set due to a more balanced strength among several co-existing goal demands. Another reason for the greater variability of set during the early trials might be the fact that cognitions relating set to the attainment of the goal are less firmly established at this time.

It appears reasonable that the efforts directed toward intentional learning diminish during the trials devoted to overlearning. This should be particularly true after *S* has correctly responded to all the syllables for several consecutive trials and perceives the task as essentially accomplished. According to the rationale developed earlier a less persistent set related to intentional learning would be expected at this stage, and consequently more incidental learning. The results substantiate the notion that the acquisition functions for intentional learning and incidental learning are different, and to some extent complementary.

The interpretations developed here are similar to those developed by Bruner, Matter, and Papanek (3) in a recent paper dealing with the dimension of breadth of learning.

Summary

The hypothesis was tested that incidental learning occurs primarily during those stages of intentional learning when motivation is relatively low. At these stages a greater variability of set is assumed as a result of a condition of near balance among several competing goal demands.

To test this hypothesis 100 *S*s were assigned to five groups and were trained in a serial memory task. The training of the various groups was interrupted after they had reached 25%, 50%, 75%, 100%, and 150% of criterion, respectively. All *S*s were tested for incidental learning immediately after their training was interrupted. It was shown that most incidental learning occurs during the very early trials, and again during the trials devoted to overlearning. It was concluded that incidental and intentional learning are to some extent complementary, with incidental learning occurring primarily when *S* is either uncertain of, or less motivated in relation to, the task set by *E*.

References

1. Bahrick, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.
2. Brown, G. H. Factors influencing incidental learning. *J. exp. Psychol.*, 1954, **47**, 163–169.
3. Bruner, J. S., Matter, J., & Papanek, L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, **62**, 1–10.
4. Postman, L., & Adams, P. A. Studies in incidental learning: III. Interserial interference. *J. exp. Psychol.*, 1956, **51**, 323–328.
5. Postman, L., Adams, P. A., & Bohm, A. M. Studies in incidental learning: V. Recall for order and associative clustering. *J. exp. Psychol.*, 1956, **51**, 334–342.

(Received August 31, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

EFFECTS OF FORCE AND AMPLITUDE CUES ON LEARNING AND PERFORMANCE IN A COMPLEX TRACKING TASK [1]

GEORGE E. BRIGGS, PAUL M. FITTS AND HARRY P. BAHRICK

The Ohio State University *Ohio Wesleyan University*

When *S* exerts a force on a control device, such as the control column of an aircraft, proprioceptive feedback information is a function of the particular spring stiffness, damping, and mass characteristics of that control (**1, 2, 7**). Of these three control loading variables, spring stiffness has received major attention, with interest directed particularly toward the relative effectiveness of force vs. amplitude cues in the use of a spring-centered control.

Weiss presents two sets of data (**10, 11**) which suggest that variation in displacement of a control (amplitude cues) can significantly affect positioning accuracy while changes in the spring constants (force cues) did not appear to so affect performance. These results were not consistent with previous data by Gibbs and Baker (**5**) which show a superiority of a pressure stick (maximum force, minimum amplitude cues) over a free control (maximum amplitude, minimum force cues) in a continuous tracking task, nor were the Weiss data in agreement with Orlansky's deductions (**9**) which indicate spring centering and nonlinear torque changes as most desirable features for control devices.

A report by Bahrick, Bennett, and Fitts (**1**) suggests a possible reason for this discrepancy: There appears to be a significant interaction between force and amplitude cues in a simple positioning task when that task involves both knowledge of results and extended practice. Thus, both force and amplitude cues can contribute to the skillful execution of simple positioning responses, and the effectiveness of either set of cues is limited by the level of the other set.

The present research attempted to determine the relative effectiveness of force and amplitude cues on performance and on learning in a complex two-dimensional tracking task. This is the second of a series of studies to assess training task effectiveness in terms of performance on a transfer task which simulates the tracking situation present in a modern interceptor aircraft.

METHOD

Apparatus.—The OSU Pilot Training Research Simulator (**6**) served as the skill task. This apparatus includes an analog computer that simulates certain characteristics of an interceptor aircraft with a radar fire-control system. The *S* views a cathode ray tube (CRT) compensatory tracking display and attempts to remove an error set in by *E* at the start of each run, by manipulating a spring-centered control column. Lateral deflections of the control

[1] This research was conducted in the Laboratory of Aviation Psychology of The Ohio State University and was supported in part by the United States Air Force under Contract No. AF 18(600)-1201 with the OSU Research Foundation, monitored by Dr. Jack A. Adams, who served as Scientific Officer for the AF Personnel and Training Research Center. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

generate a voltage, *X*, which is subjected to two stages of integration, the first representing rate of roll of the aircraft, the second simulating rate of change in heading. Fore and aft deflections generate another voltage, *Y*, which is subjected to an exponential oscillatory function plus one stage of integration, the combination representing the underdamped response of an aircraft in the pitch dimension.

The initial system input is a step function and one may consider the target "course" as being of zero relative velocity. The *S*'s goal is to remove the initial error by returning the target dot to the center of the CRT, the appropriate area being shown as a circle of ¼-in. diameter on the display. Movements of the target on the CRT result from two effects: (*a*) *S*'s control movements, which are subjected to the transformations cited above, and (*b*) a constantly increasing magnification of the error, which is achieved by multiplying the *X* and *Y* error components by $1/T$, where *T* represents time-to-go in a trial and is continuously variable from 40, at the beginning, to 1 at the end of each trial. More detailed descriptions of the apparatus appear elsewhere (**3, 6**).

Subjects.—The *S*s were 68 freshmen and sophomore male students. All were volunteers from the Air Command Squadron of the Air Force ROTC unit at The Ohio State University, none of whom had received prior flight training. Each *S* was paid for five experimental sessions of approximately 45 min. each.

Procedure.—During the first session *E* gave *S* a demonstration of the task requirements, following which each *S* received 10 40-sec. trials of tracking with 20-sec. rest between trials and 5-min. rest between blocks of five trials. The *S*s served in pairs; one *S* tracked for five trials while his companion observed, then the second *S* tracked for five trials, and so on. During the following four sessions each *S* received a total of 20 40-sec. trials per session. A 24-hr. interval separated each of the five sessions. There were a total of 60 training trials and 30 transfer trials. Transfer was made halfway through Session 4.

The *S*s were assigned at random to one control and three experimental groups of 17 *S*s each. A systematic variation of the force and amplitude cues present in the control column defined groups. A summary of these conditions during the 60 training trials for lateral control deflections is given in Table 1. The vertical or pitch dimension of the tracking task was modified in an identical manner. Group 1 served as the control group and these *S*s trained with a control column involving force and amplitude cues approximating those found in an interceptor aircraft at 500 kts. and at 35,000 ft. Columns 4 and 5 present physical specifications of the force cues while Column 6 indicates the relative level of "aerodynamic" or control sensitivity, S_e, of the tracking system. As shown in Columns 2 and 3, the "normal" control cues are designated as "1." These normal cues were those present in the control column used in a previous study (**3**).

TABLE 1

FORCE-DISPLACEMENT CONDITIONS FOR ROLL MANEUVERS

Group	Force	Amplitude	Centering Force (lb.)	Lateral Displacement Force (lb./deg.)	Sensitivity*
1	1	1	2	1	10
2	¼	1	1	½	10
3	1	¼	6	2	40
4	¼	¼	2	1	40

* These are arbitrary values for computer sensitivity (S_e) or "aerodynamic sensitivity" for roll (deg./sec.^{2} per degree of control displacement). Exact specifications appear elsewhere (**6**).

Group 2 received their training trials with a control in which the force cues (relative torque change) were attenuated by a factor of 4, compared to the "normal" stick. This was accomplished by reducing the spring constant, as shown in Table 1. Computer sensitivity to a given amount of control column movement was unchanged, leaving the amplitude cues at the same level as for Group 1. Therefore, *S*s in Group 2 were required to displace the control to the same extent as Group 1 in order to achieve the same system response, but at those amplitudes Group 2 experienced only one-fourth the force feedback in terms of relative torque change.

For Group 3 the spring constant was increased to four times that for Group 1 while computer sensitivity also was increased to four times "normal." These *S*s now deflected the control only one-fourth the amplitude required of Group 1 for the same system response, but at that amplitude Group 3 experienced the same total force feedback as that received by the control group at four times the amplitude. Finally, Group 4 tracked with a control column in which both force and amplitude cues were reduced as compared to Group 1. For Group 4 the spring constant present in the control column for Group 1 was used, but the computer sensitivity was increased by a factor of four over that experienced by the control group. Therefore, these *S*s achieved the same system response by deflecting the control one-fourth the amplitude needed by Group 1, but at that amplitude

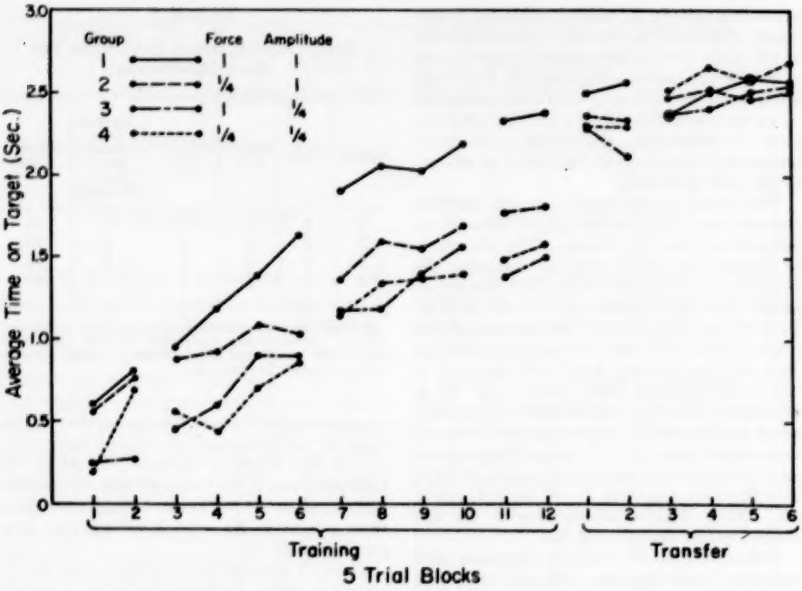

FIG. 1. Training and transfer performance levels for all groups over the five experimental days. Transfer was affected during Session 4. Maximum possible score was 3 sec.

Group 4 experienced only one-fourth the force change received by Group 1.

During the 30 transfer trials, Groups 2, 3, and 4 tracked with the same control column force and amplitude cues as the control group, Group 1. Thus, the experiment provides two sets of related data: (*a*) during the training trials it is possible to assess the influence of changes in force and amplitude cues on tracking performance, and (*b*) during the transfer trials it is possible to determine the influence of force and amplitude cues on the learning of this skill task, independent of the specific effects of these cues on performance during training.

Tracking proficiency was measured in terms of time on target simultaneously in X and Y. The scoring area was a square ¼ in. × ¼ in. at the center of the CRT; this approximates the target circle which was ¼ in. in diameter. This score was cumulated over the final 3 sec. of each tracking trial. A previous study (**3**) indicated that particular period of scoring as satisfactory for this skill task.

RESULTS

The training and transfer data for all groups over the five sessions of the experiment are presented graphically in Fig. 1. Each point is the mean performance of 17 *S*s as determined by the time-on-target scores averaged for five-trial blocks. There were two blocks of trials on Session 1 and four blocks on each of the remaining four sessions. Transfer was made midway through Session 4.

Training performance.—During the 60 training trials, Group 1 shows a consistent superiority over the three experimental groups, with Group 2 enjoying a slight superiority over Groups 3 and 4. These latter two groups exhibit no consistent superiority one to another. It appears, then, that reduction in force and/or amplitude cues can influence tracking performance.

In order to assess the relative importance of force vs. amplitude cues on performance, a simple 2 × 2

TABLE 2

ANALYSIS OF VARIANCE FOR TRAINING BLOCKS 11 PLUS 12

Source	*df*	*MS*	*F*
Force (F)	1	509.23	2.49
Amplitude (A)	1	2639.59	12.90*
F × A	1	972.86	4.75*
Residual	64	204.66	

* $P < .05$.

analysis of variance was performed on the data for the final 10 training trials. For this analysis the data for Blocks 11 and 12 were averaged for each *S* in each group. The results of this analysis are presented in Table 2. The lack of statistical significance ($P > .05$) for the Force variable is in agreement with previous analyses by Weiss (**10, 11**), and considered alone, this would suggest that only Amplitude was the important ($P < .01$) variable in this tracking task. However, the statistical significance of the Force × Amplitude interaction ($P < .05$) qualifies these conclusions. This interaction implies that both force and amplitude cues influenced tracking proficiency here, and Table 3 shows this quite clearly.

In Table 3 the time-on-target scores were averaged for groups and then arranged according to the levels of the force and amplitude variables used in this experiment. The Duncan Multiple Comparisons Test (**4**) was applied to these means, and the results show that Group 1 performance (2.36 sec. on target) differed at the .01 level of confidence from the performance levels of Groups 2, 3, and 4, the latter three groups exhibiting no significant differences one from another.

TABLE 3

AVERAGE TIME-ON-TARGET SCORE OVER TRAINING BLOCKS 11 AND 12
(Maximum possible score = 3.0 sec.)

Relative Amplitude	Relative Force	
	1	¼
1	2.36	1.80
¼	1.45	1.54

Thus, it appears that changes in amplitude cues significantly affected performance but only when force cues were high (had a value of "1") not when force cues were reduced (had a value of "¼"), and likewise changes in force cues significantly affected performance when amplitude cues were high, but not when these latter cues were at a low level. This significant interaction confirms the expectation based on the data of Bahrick, Bennett and Fitts (**1**): force and amplitude interact to influence the levels of performance in skill tasks which involve controls with spring loading.

It is to be noted, however, that the particular pattern of differences found here to define the Force × Amplitude interaction differed from that pattern noted by Bahrick, Bennett, and Fitts. In that earlier study, force cues did not appear to differentiate performance at high levels of the amplitude variable (70° movements of a unidimensional control) but did have a significant effect on lower levels (35° and 17.5° displacements). It would appear that a particular interaction pattern is dependent primarily upon the slope of the function relating spring stiffness to control displacement and the portion of that function present in a given experiment. Therefore, the statistical significance of the interaction, per se, is the important finding at this time. Considerable additional research is needed to define the nature of this interaction more completely.

Transfer performance.—By visual inspection of Fig. 1 it can be seen that a high degree of positive transfer occurred for each of the three experimental groups. The significance of this observation was substantiated by

TABLE 4

ANALYSIS OF VARIANCE FOR TRAINING BLOCKS 11 PLUS 12 AND TRANSFER BLOCKS 1 PLUS 2

Source	*df*	*MS*	*F*
Groups (G)	3	2319.02	3.72*
Ss/G	64	623.72	
Training-Transfer (T-T)	1	11 492.52	113.11*
T-T × G	3	707.44	6.96*
T-T × *Ss*/G	64	101.61	
Blocks (B)	1	5.95	—
B × G	3	9.20	—
B × *Ss*/G	64	90.45	
B × T-T	1	111.90	2.06
B × T-T × G	3	24.84	—
B × T-T × *Ss*/G	64	54.28	

* $P < .05$.

an analysis of variance for the data of Training Blocks 11 and 12 and for Transfer Blocks 1 and 2. The results of this analysis are presented in Table 4. The Training-Transfer comparison shows a highly significant effect. A more analytic expression of this transfer is justified by the significance of the Training-Transfer × Groups interaction, i.e., application of the Duncan Test to the eight mean scores defining this interaction gives the results listed in Table 5. Those means shown within brackets do not differ significantly, one from another, while those means not within a bracket do differ at the .05 level. As shown earlier in Table 3, Group 1 is significantly superior to the three experimental groups at the end of the training sessions. During the transfer trials, however, this difference is diminished markedly. It is concluded, then, that near 100% transfer was found for each of the three experimental groups.

TABLE 5

AVERAGE TIME-ON-TARGET FOR TRAINING BLOCKS 11 PLUS 12 AND TRANSFER BLOCKS 1 PLUS 2 (Maximum possible score = 3.0 sec.)

Group	Training	Transfer
1	2.36	2.54
2	1.80	2.36
3	1.45	2.21
4	1.54	2.31

A convenient index to express this high degree of positive transfer is $\frac{E_T - C_1}{C_T - C_1} \times 100$, where C_1 is the initial performance level of Group 1 (.6 sec.), E_T is the transfer performance level of an experimental group (2.36, 2.21, and 2.31 sec. for Groups 2, 3, and 4, respectively), and C_T is the performance of the control group, Group 1, at the time of transfer (2.54 sec.). These indices are 90%, 83%, and 88% for Groups 2, 3, and 4, respectively.

DISCUSSION

The data presented in Fig. 1 indicate that whereas the particular combination of force and amplitude cues present in a control column can set rather definite limits on tracking performance level, there apparently is little retardation in the learning of the skill task as compared to learning under either more or less "favorable" control characteristics.

Transfer effects.—This latter finding is in accord with an hypothesis stated in an earlier report by the authors (**3**). There it was proposed that practice with any particular range of response amplitudes should involve near 100% transfer to a task permitting a different range of response amplitudes provided (*a*) no important system transformations were changed from the training to transfer tracking systems, and (*b*) the responses practiced during training were not qualitatively different from those required during transfer.

The first restriction was met in the present study as the system transforma-

tions for both elevation and azimuth corrections were constant from training to transfer tasks (only the values of the constants for the mathematical operations were changed). The second restriction was met in that Groups 3 and 4 trained on a system requiring control displacements of only lesser amplitudes than those of Group 1—the training and transfer responses were quantitatively not qualitatively different. Group 2 utilized a control permitting substantially the same range of displacement amplitudes as for Group 1.

Training effects.—It is clear from Fig. 1 that 4-to-1 changes in amplitude cues produced a greater effect on tracking performance (Group 3 vs. Group 1) than a comparable ratio of change in force cues (Group 2 vs. Group 1). This is in agreement with previous data (**1, 10, 11**). It is the purpose of this section to show that the particular order of proficiency levels, found here for the four groups during training, is a logical order and one consistent with previous information on force discrimination.

Bahrick, Bennett, and Fitts (**1**) propose, on the basis of data from their study and that of Jenkins (**8**), that the ratio $\frac{\Delta F}{F(\Delta D)}$ provides an index predictive of accuracy in positioning a spring-centered control. Here, F is the force required to achieve a specified control displacement, and ΔF is the force change associated with some given displacement, ΔD. The empirical data suggest that as this ratio is increased for a constant displacement, positioning accuracy increases.

An interpretation of this ratio is facilitated by considering the force-displacement plot shown in Fig. 2 for the four spring-amplitude conditions used in the present study. These functions were determined empirically for the springs involved in lateral control deflections; similar functions hold for all directions of stick displacement. For purposes of discussion we will consider the spring conditions at 8° displacement of the control column for Groups 1 and

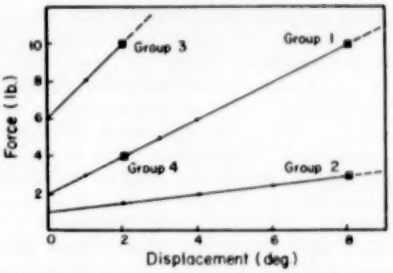

FIG. 2. The force-displacement relations encountered by the groups during training. Points indicated refer to typical displacements during terminal phases of each trial.

2 and at 2° for Groups 3 and 4. These values of ΔD are typical of the amplitudes of control deflection utilized by S during the latter portion of a tracking trial.

At these two values of ΔD, ΔF is 8, 2, 4, and 2 lb., and F is 10, 3, 10, and 4 lb. for Groups 1 through 4, respectively. Thus, the ratios are .1, .08, .20, and .25 for Groups 1 through 4. For a given displacement, $\Delta D = 8°$, the ratios .1 and .08 predict that Group 1 will be superior to Group 2, and for $\Delta D = 2°$ the ratios .20 and .25 predict that Group 4 will be superior to Group 3 in tracking performance. According to Bahrick, Bennett, and Fitts (**1**), however, prediction of the utility of force cues in improving the accuracy of positioning movements depends not only upon this ratio (relative torque change per unit distance), but also upon the absolute force change involved in control deflections.

When S applies force to a control, Jenkins (**8**) found that the relative DL decreases significantly from 1 to 10 lb., so that within this range the larger absolute values of force change are much more useful than smaller values, given a particular ratio of relative torque change. Thus, in comparing the conditions for Groups 1 and 2, Group 1 encountered the larger absolute changes in force, and so superior performance for this group over that for Group 2 would again be predicted. In comparing the force-dis-

placement conditions for Groups 3 and 4, however, Group 3 experienced the larger absolute change of force, while Group 4 encountered the larger relative change per unit of displacement, as noted above. On the basis of these counteracting conditions, then, a clear-cut prediction of the relative tracking accuracies of Groups 3 and 4 is not possible, and the present finding of no significant differences between these two groups during training (Table 5) is not unexpected.

Summary

Four groups of male *Ss* served for 60 training and 30 transfer trials in a study designed to assess the influence of force and amplitude cues from a control column on learning and performance in a two-dimensional compensatory tracking task. Group 1 trained with a control in which the force-displacement relation was known (**3**) to permit relatively good performance with the tracking system. Group 2 trained on a control which involved one-fourth the total force change, but which permitted displacement of the same amplitudes as for Group 1. The control column used by Group 3 involved the same total force feedback as that for Group 1, but amplitudes of displacement were reduced to one-fourth. Finally, Group 4 trained on a control which had both force and amplitude cues reduced to one-fourth those present for Group 1. After the 60 training trials all groups practiced for an additional 30 trials on the control column used by Group 1.

Over the 60 training trials Group 1 was significantly superior to the three experimental groups with Group 2 being slightly superior to Groups 3 and 4. On the first block of transfer trials there were no significant differences among the four groups, thus indicating near 100% transfer for the three experimental groups.

Analysis of the scores during the final training trials revealed that both force and amplitude cues significantly affected performance, amplitude cues apparently exerting the greater influence. These performance results were shown to be predictable from the ratio $\frac{\Delta F}{F(\Delta D)}$, where ΔF is a force change associated with a given displacement change, ΔD, and F is the terminal force required to attain the displacement.

References

1. Bahrick, H. P., Bennett, W. F., & Fitts, P. M. Accuracy of positioning responses as a function of spring loading in a control. *J. exp. Psychol.*, 1955, **49,** 437–444.
2. Bahrick, H. P., Fitts, P. M., & Schneider, R. Reproduction of simple movements as a function of factors influencing proprioceptive feedback. *J. exp. Psychol.*, 1955, **49,** 445–454.
3. Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Learning and performance in a complex tracking task as a function of visual noise. *J. exp. Psychol.*, 1957, **53,** 379–387.
4. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2,** 171–189.
5. Gibbs, C. B., & Baker, J. C. Free-moving versus fixed control levers in a manual tracking task. In A. Tustin (Ed.,) *Automatic and manual control.* New York: Academic Press, 1952.
6. Harter, G. A., & Fitts, P. M. The functional simulation of complex systems by means of an analog computer, with the F-86D, E-4 system as a specific example. AFPTRC, *Res. Bull.* (In press.)
7. Howland, D., & Noble, M. E. The effect of physical constants of a control on tracking performance. *J. exp. Psychol.*, 1953, **46,** 353–360.
8. Jenkins, W. O. The discrimination and reproduction of motor adjustments with various types of aircraft controls. *Amer. J. Psychol.*, 1947, **60,** 397–406.
9. Orlansky, J. Psychological aspects of stick and rudder controls in aircraft. *Aero. Engng Rev.*, 1949, **8,** 1–10.
10. Weiss, B. The role of proprioceptive feedback in positioning responses. *J. exp. Psychol.*, 1954, **47,** 215–224.
11. Weiss, B. Movement error, pressure variation, and the range effect. *J. exp. Psychol.*, 1955, **50,** 191–196.

(Received September 14, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

EFFECT OF VARIATION IN ASSOCIATIVE FREQUENCY OF STIMULUS AND RESPONSE MEMBERS ON PAIRED-ASSOCIATE LEARNING[1]

GEORGE MANDLER AND ENID H. CAMPBELL

Harvard University

The frequency of associations elicited by verbal materials has been shown by several investigators to exert a significant effect upon verbal learning. Noble's index of meaningfulness (*m*), which is determined by the number of associations elicited during a 1-min. period, has been shown to be related to the rate of serial learning (**7**) and the rate of acquisition of paired associates (**2, 8**). Similarly, paired-associate learning of nonsense-syllable pairs is related to the index of associative frequency, determined by the number of associations elicited during a 30-sec. period (**5**). These studies indicate that paired-associate acquisition is a positive function of the number of associations elicited by the stimulus and response members of the paired associates.

This paper will present two experiments in which the associative frequencies of the stimulus and response members of nonsense-syllable paired associates are varied independently. Noble has indicated that the associative frequency variable can be related to variations in meaningfulness. Thus, meaningful words show a greater frequency of associations than nonsense words (**6**). In an unpublished study, Sheffield (**9**) has investigated the effect of two levels of meaningfulness, three-letter words and three-letter nonsense syllables, on rate of acquisition in paired-associate learning. Two of Sheffield's findings are relevant to the present study. The first was that learning was greatly facilitated by the presence of a meaningful word in the response position, while in the stimulus position the trend toward facilitation by the presence of a meaningful word was smaller and nonsignificant. The other result indicated that prior familiarization with nonsense syllables has a definite facilitative effect when the syllable is used in the response position, but a transitory negative effect when it is used in the stimulus position.

The two experiments presented below are concerned with the effect on paired-associate learning when both stimulus and response members are varied on three levels (high, medium, and low) of associative frequency, and with the effect of prior associative activity of *S*, particularly when he is required to produce associations to the syllables used in the learning task.

METHOD

Design.—Experiments I and II both investigated the acquisition of paired-associate nonsense syllables. They differ in the lengths of lists and dependent measures of acquisition utilized. All syllables were selected from the list of 100 consonant-vowel-consonant syllables whose associate frequency values have been published elsewhere (**4**). The values of these 100 syllables vary from 2.9 to 5.3 associations elicited in a 30-sec. interval. High, medium, and low associative-value syllables were obtained

[1] This research was supported by a grant from the Laboratory of Social Relations, Harvard University. The authors wish to thank Dr. Robert P. Abelson for his advice on the statistical treatment of the data, and Drs. F. D. Sheffield and C. E. Noble for other helpful suggestions.

by starting the selection procedure from the two ends of the distribution and on both sides of the median. Syllables were included in any one list such that for any one pair the same letter did not appear in both the stimulus and response member, and that no pair of adjoining letters appeared in more than one syllable. In addition, in Exp. I the initial letter of a stimulus syllable did not appear in the initial position in any other stimulus syllable.

In Exp. I three lists (A, B, and C) of six pairs were used. The resultant 18 pairs consisted of two instances of each of the possible nine combinations of high, medium, and low stimulus members (S_h, S_m, S_l) and high, medium, and low response members (R_h, R_m, R_l). Thus, List A contained two pairs each representing the S_h-R_l, S_m-R_m, and S_l-R_h condition. List B contained the S_h-R_m, S_m-R_h, and the S_l-R_l conditions. List C represented the S_h-R_h, S_m-R_l, and S_l-R_m conditions. One half of all the *S*s given any one list were presented with an association task prior to the learning task, the other half were given the association task subsequent to learning. The association task consisted of 20 nonsense syllables different from those used in the learning task, and *S*s wrote down all the associations elicited by these syllables during a 30-sec. interval.

In Exp. II one list of 18 pairs was used, two pairs for each possible combination of high, medium, and low stimulus and response members. The *S*s were randomly assigned to three groups. One group was given, prior to the learning task, an association task which used the 36 syllables used in the learning task. A second group was given, prior to learning, 36 syllables different from those used in the learning task. The third group was given no prior association task.

Table 1 shows the mean associative frequency values of the stimulus and response members in the two experiments.

TABLE 1

MEAN ASSOCIATIVE FREQUENCY OF STIMULUS AND RESPONSE TERMS

Exp.	Stimulus			Response		
	High	Med.	Low	High	Med.	Low
I	4.9	4.0	3.6	4.7	4.1	3.6
II	4.9	4.0	3.5	4.9	4.0	3.3

Procedure.—The lists were presented by means of a Hull-type memory drum. The stimulus member was exposed for 1.5 sec., followed by a .5 sec. interval during which both shutters were closed, and a 1.5 sec. exposure of the response member, with another .5 sec. interval during which both shutters were closed. Correct responses were scored as anticipations of the response member prior to its exposure. The lists in Exp. I were presented in six different randomized orders, which sequence was repeated starting with the seventh presentation, and so forth. In Exp. II three randomized orders of the list were repeated in a similar manner. The *S*s in Exp. I continued on the learning task until they achieved at least one correct response for each pair in the list. In Exp. II, where due to the length of the list a criterion of correct responses was not feasible, all *S*s were given 30 trials.

In the association task *S* was successively presented with the nonsense syllables printed on 3 × 5-in. filing cards. Each syllable was presented for 30 sec., during which time *S* wrote down single word associations.

Subjects and instructions.—A total of 66 volunteer paid undergraduate *S*s were used.

TABLE 2

MEANS AND *SD*'S OF TRIALS TO FIRST CORRECT RESPONSE IN EXP. I

Response Terms	Stimulus Terms							
	S_h		S_m		S_l		Total	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
R_h	**9.6**	**9.1**	*9.7*	*1.3*	15.5	12.3	11.6	9.2
R_m	*10.1*	*6.8*	19.0	10.8	**10.4**	**6.1**	13.2	9.0
R_l	19.0	13.5	**17.3**	**13.0**	*10.8*	*5.6*	15.7	11.7
Total	12.9	10.9	15.3	10.5	12.2	8.6		

Note.—Data from List A is indicated in normal Arabic numerals, from List B in *italics*, and from List C in **bold face**. The analysis of variance in Table 4 is based on a logarithmic transformation of these data.

TABLE 3

NUMBER OF CORRECT RESPONSES IN EXP. II

Response Terms	Stimulus Terms			
	S_h	S_m	S_l	Total
R_h	17.9	10.6	11.6	13.3
R_m	11.8	13.4	10.1	11.8
R_l	13.8	8.2	7.5	9.8
Total	14.5	10.7	9.8	

Note:—The analysis of variance in Table 4 was based on the frequency of correct response to *two* paired associates representing each S-R combination. In this table these values were halved to present mean frequency for single paired associates.

Thirty-six *Ss* participated in Exp. I, 12 *Ss* for each of the three lists; and 18 *Ss* in each of the two association conditions. Thirty *Ss* participated in Exp. II, 10 *Ss* in each of the three association conditions. All *Ss* were given standard paired-associate instructions, and association instructions similar to those used in the standardization study (**4**). The connection between the two tasks was not explained to *Ss*.

RESULTS

Tables 2 and 3 show mean values of the learning scores for Exp. I and II. They indicate the mean value of all pairs in the particular stimulus or response category, e.g., in Exp. I the six pairs with a high frequency stimulus member showed an average first correct response on Trial 12.9. Table 4 summarizes the analyses of both experiments. In Exp. I the learning variable used was the number of the trial in which *S* gave the first correct response for a particular nonsense syllable pair. For purposes of analysis the logarithm of this measure was used. In Exp. II, the dependent measure was the number of correct responses given in 30 trials (presentations). Since every one of the nine possible stimulus-response combinations was represented by two pairs, the values for these parallel pairs were combined. Learning effects were estimated by using successive blocks of six trials (T).

The analyses of variance show agreement only on the two major experimental variables, variation of stimulus and of response members.

TABLE 4

SUMMARY OF ANALYSES OF VARIANCE OF EXP. I AND II

Source	Experiment I*					Experiment II				
	df	Error Term	*MS*	*F*	*P*	*df*	Error Term	*MS*	*F*	*P*
1. Groups (G)	1	10	.085	<1		2	10	15.0	<1	
2. Stimuli (S)	2	11	.195	6.7	.01	2	12	452.0	25.7	.001
3. Responses (R)	2	11	.307	10.5	.001	2	13	223.5	12.6	.001
4. S × R	2	11	.073	2.5	.10	4	14	134.5	6.0	.01
5. S × G	2	11	.055	1.9	>.10	4	12	15.2	<1	
6. R × G	2	11	.093	3.2	.05	4	13	7.5	<1	
7. S × R × G	2	11	.012	<1		8	14	15.4	<1	
8. Trials (T)						4	15	2329.2	161.8	.001
9. T × G						8	15	6.6	<1	
10. Between *Ss* Error	30		.141			27		220.4		
11. Within *Ss* Error	155		.029							
12. *Ss* × S						54		17.6		
13. *Ss* × R						54		17.7		
14. *Ss* × S × R						108		22.5		
15. *Ss* × T						108		14.4		

* In Exp. I lists are confounded with Stimulus × Response interactions. All effects are reported with the confounding effects of lists removed. It should be noted, however, that variance due to lists was significant at the .01 level. Gourlay (1) has recently pointed out that in this latin-square type design, such a finding might indicate that the variance associated with the Stimuli, Responses, and Stimuli × Responses effects is inflated.

These two sources of variation are significant in both experiments at better than the .01 level. Tables 2 and 3, however, indicate that while response variation in both cases shows increasing facilitation as associative frequency is increased from low to medium to high, stimulus variation shows this effect only in Exp. II. In Exp. I, the S_h members are learned significantly faster than the S_m members, but not significantly better than the S_l members.

Two other significant findings are shown in Table 4. Neither of these is replicated in both experiments. In Exp. I, there is a barely significant interaction between Response variation and Association groups. This significant interaction is associated with faster learning of the R_m and R_l pairs when learning is preceded by the association task, with no such difference for the R_h pairs.

The other finding, in Exp. II, shows a significant Stimulus-Response interaction. This interaction effect is apparently primarily effected by three of the stimulus-response combinations. It is apparent from Table 3 that the S_h-R_l pairing produced unusually fast acquisition, particularly when compared with the combination of the R_l member with S_m and S_l members.

This rapid acquisition of S_h-R_l pairs is in contrast with an incidental datum from Kimble and Dufort (**2**). These investigators found that lists using Noble's disyllables and three-letter words were more easily learned when the less meaningful disyllables were in the stimulus position and the three-letter words in the response position than when this relationship was reversed. Similarly Sheffield's data show that pairs with nonsense syllables as stimuli and meaningful words as responses are learned faster than pairs with meaningful words in the stimulus position and nonsense syllables in the response position (**7**). These findings would imply in the present study that the S_l-R_h pairs should be learned faster than the S_h-R_l pairs; however, the opposite is the case.[2] One possible explanation of these discrepant findings is that in both the Kimble and Dufort and the Sheffield studies, the *S*s had to learn "how to respond" when the more difficult (less meaningful) items were in the response position. Both the majority of the disyllables used by Kimble and Dufort and the nonsense syllables used by Sheffield required *S* to learn some novel combinations of sounds, i.e., introduce a selective effect of response learning (**3**). This requirement might increase acquisition time for nonsense responses as against meaningful, previously practiced, responses. In the present study this learning of the response was required in all conditions. Thus the saving from previous experience would not be a factor which would adequately facilitate the acquisition of more meaningful responses.

The other two values contributing to the significant S × R interaction are the S_m-R_h pairing which shows a somewhat lower rate of acquisition than expected, and the S_m-R_m pairing which shows a somewhat high rate of acquisition. In this connection it should be pointed out that the S_h-R_h, S_m-R_m, and S_l-R_l pairings show the expected decreasing rate of learning. This is consistent with the previous research (**2, 5, 8**).

None of the results indicates any simple effects on learning as a function of previous exposure to the learning material. Neither the presentation of identical material in a prior association task, nor an association task with different materials shows any, even suggestive, effect on rate of learning in these two experiments.

The present results indicate that associative frequency is an important

[2] It should be noted that in Exp. I, where the Stimulus-Response interaction shows a level of significance of only .10, the data are in the direction of the Kimble and Dufort findings.

factor in paired-associative learning. This variation is effective in both the stimulus and the response members of nonsense syllable pairs. However, only in the response member was it a consistent function of associative frequency in both experiments. Variation of the stimulus member showed a similar effect in Exp. II, but in Exp. I no such simple function appears. This may have been partly due to confounding effects in the design of Exp. I. This discrepancy is particularly important in view of Sheffield's previous findings of a consistent effect of the variation of meaningfulness on the response side, but not on the stimulus side. However, it should be noted that Sheffield's criterion of meaningfulness was the presence of the word in a dictionary. Thus his meaningful items had a high degree of familiarity which is absent for the syllables in the present studies.

The failure to find any consistent significant effects associated with prior associative activity on the part of the *Ss* must be contrasted with Sheffield's findings of a significant facilitative effect of extensive familiarization when a nonsense syllable is used in the response position. The 30-sec. association period may have been quite insufficient to provide *S* with adequate experience with the nonsense syllable.

Summary

Two experiments were performed in which *Ss* learned nonsense-syllable paired associates. The pairs in both experiments represented three degrees of associative frequency (high, medium, and low) in both the stimulus and the response members. In addition, different groups of *Ss* were given different association tasks prior to the learning task. The main findings were:

1. Variation in associative frequency of the response member showed increasing facilitation in learning as a function of increasing degrees of associative frequency.

2. While variation of associative frequency of the stimulus members also showed a significant effect on acquisition, this effect was not consistently in the direction of a simple association between associative frequency and learning.

3. The presence or absence of prior association tasks did not significantly affect acquisition in the learning task.

References

1. Gourlay, N. *F*-test bias for experimental design of the latin square type. *Psychometrika*, 1955, **20**, 273–287.

2. Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.

3. Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

4. Mandler, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. *Amer. J. Psychol.*, 1955, **68**, 662–665.

5. Mandler, G., & Huttenlocher, J. The relationship between associative frequency, associative ability and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.

6. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

7. Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; Erratum, 1952, **44**, 465.

8. Noble, C. E., & McNeely, D. A. The role of meaningfulness (*m*) in paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 16–23.

9. Sheffield, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctor's dissertation, Yale Univer., 1946.

(Received September 28, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

STUDIES OF DISTRIBUTED PRACTICE: XVII. INTERLIST INTERFERENCE AND THE RETENTION OF PAIRED CONSONANT SYLLABLES [1,2]

BENTON J. UNDERWOOD AND JACK RICHARDSON

Northwestern University

A previous study (**13**) demonstrated that retention of serial consonant syllable lists was inversely related to the number of lists learned prior to the acquisition of the list tested for retention. Furthermore, in that study there was an interaction between intertrial interval and the number of previous lists learned, namely, massed practice of the first-learned list gave better retention than did distributed practice but just the reverse was true for the seventh list in the series. On the assumption that as number of previous lists increase interlist interference increases, the results indicated that massed practice facilitates retention when interlist interference is low and distributed practice facilitates when this interference is high.

It is commonly found that retention of paired-associate lists is better following distributed practice than following massed practice (**4, 7, 8**) but the differences are not usually statistically significant. In these studies *S* served in counterbalanced conditions and learned a total of four lists so there was some possibility for interlist interference. In comparable experiments with serial lists (**5, 6, 10**) massed practice was followed by better retention than was distributed practice. Other analyses (**11**) suggest that interlist interference effects are greater for paired-associate lists than for serial lists.

The above facts suggest the possibility that whether or not distributed practice facilitates retention is due entirely to amount of interlist interference and that the type of learning involved is not a critical variable. Moreover, these facts suggest that the "necessary" amount of interlist interference needed to produce significantly better retention following distributed practice of paired-associate lists would be present with fewer lists than in the case of serial learning. To obtain data bearing on this issue the amount of interlist interference is varied in the present experiment by varying the number of lists learned previous to the learning of a list by distributed practice. At one extreme there are six previous lists (thus duplicating the serial experiment), at the other, one. In addition, the lists have been constructed in the same way as the serial lists so that interlist formal similarity is high.

METHOD

Lists.—Each of seven paired-associate lists was constructed of four pairs of consonant syllables taken from Melton (**2**). All syllables were between 50% and 75% association value as calibrated by Witmer (**14**). The 56 syllables contained a total of nine different consonants. Each of the nine consonants appeared from one to four times in each list. The lists were presented on a Hull-type memory drum at a 2:2-sec. rate and learning was by the anticipation method

[1] R. W. Schultz supervised the gathering of the data.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

in which *S* spelled the syllable. Six orders of presentation were used to prevent serial learning.

Design.—A total of 336 undergraduates served as *S*s. These were divided into six groups of 56 each. Two of these groups learned a total of 7 lists, two learned 4 lists, and two learned 2 lists. For each number of lists there was a massed learning group and a distributed learning group. The procedure for the 7-list groups parallels that of the previous serial experiment. The distributed group learned the first list with a 30-sec. intertrial interval. Twenty-four hours later this list was recalled and relearned and then two additional lists were learned by massed practice (4-sec. intertrial interval). On the third day three more lists were learned by massed practice. On the fourth day a single list was learned with a 60-sec. intertrial interval, and 24-hr. later this list was recalled and relearned. The corresponding massed group went through exactly the same schedule except that all lists were learned by massed practice.

The 4-list groups proceeded the same as the 7-list groups through the first two days. On the third day a single list was learned; this list was then recalled and relearned 24 hr. later. One of the groups learned all four lists by massed practice, the other learned the first list with a 30-sec. intertrial interval and the fourth with a 60-sec. interval, with the two "filler" lists being learned by massed practice.

The 2-list groups learned and recalled the first list as did the other groups. However, on the second day, after relearning the first list, a second list was learned which was recalled 24 hr. later. One group learned both lists by massed practice, the other with the 30-sec. and 60-sec. intertrial intervals for the first and second list, respectively.

The six groups will be designated by a number and a letter, the number designating the total number of lists learned and the letter whether the first and last lists were learned by massed or distributed practice. Thus, the 4-D group learned four lists and the learning of the first and fourth list was with a 30-sec. and 60-sec. intertrial interval, respectively. The six groups were 2-D, 2-M, 4-D, 4-M, 7-D, and 7-M.

The groups which learned different numbers of lists did not learn the same lists. The four lists were selected from the seven and the two from the four. Within each number of lists the different lists were systematically varied so each list appeared equally often at each stage of practice. Standard instructions for paired-associate anticipation learning were given prior to the learning of the first list and at relearning it was emphasized that *S* should anticipate as many syllables as possible on the very first trial. The intertrial intervals longer than 4 sec. were filled with symbol cancellation (**5**). All learning and relearning was to a criterion of one perfect trial. Relearning was by massed practice for all groups for all lists.

Results

Learning.—The mean number of trials required to learn each list to the criterion of one perfect trial is shown in Fig. 1. The graph is drawn as if it represents only the two groups learning seven lists. While one curve is labeled "massed" and the other "distributed" it should be remembered that this difference in conditions holds only for Lists 1 and 7; Lists 2 through 6 were learned by massed practice. The points in Fig. 1 are based on all data available for a given list in the series; thus *N* decreases as the 2- and 4-list groups drop out. For the massed curve the *N* is 168, 168, 112, 112, 56, 56, and 56 for Lists 1 through 7, respectively. The corresponding *N*'s for the distributed curve are 168, 112, 112, 56, 56, 56, and 56. The learning of List 2 by Group 2-D and List 4 by Group 4-D are shown by separate points.

It may be noted first that the change in performance within days and between days for Lists 2–6 show a remarkable correspondence with the results from the previous serial experiment (**13,** Fig. 1). There is no

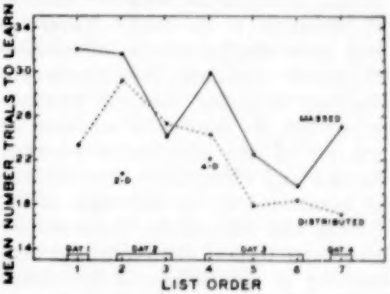

Fig. 1. Mean number of trials to learn seven successive lists of paired consonant syllables.

way in the present data of determining how much of these changes can be attributed to each of various phenomena, such as warm-up, practice effects, and interference. As far as learning is concerned, the major interest is in the apparent consistently faster learning with distributed practice than with massed practice. The mean number of trials required to learn Lists 1, 2, 4, and 7 by distributed practice is in each case less than the number of trials required to learn the corresponding massed list. However, it is possible that these differences may result, at least in part, from the selection of *Ss* by the conditions of the experiment. Although the lists had only four pairs of syllables, they were very difficult to learn; 135 *Ss* were dropped from the experiment, of which 60 were dropped for failure to learn within a 50-min. period. The remainder was lost for miscellaneous reasons such as failure to return for recall. The number of *Ss* dropped increased as the total number of lists to be learned increased and was greater for the distributed conditions than for the massed. Group 2-M had the fewest drops, Group 7-D the most.

The present experiment was designed to test the interaction between number of previous lists and intertrial interval on retention and equality of groups is not essential for the measure of retention to be used. Nevertheless, since the facilitation of learning of paired associates by distributed practice is rather rare, it becomes important to determine whether or not all of the differences between massed and distributed learning can be attributed to the selection of *Ss*. Within the experiment there are at least three tests of the hypothesis that learning is *not* facilitated by distribution of practice and each of these is biased in favor of the hypothesis.

1. Two groups, 4-D and 7-D, learned List 1 by distributed practice and List 2 by massed practice. Both practice effects and warm-up would tend to make List 2 learned more rapidly if distributed practice did not facilitate the learning of the first list. Quite the opposite is true. The mean number of trials required to learn List 1 (distributed practice) by the 4-D and 7-D groups combined was 23.12 ± 1.00, and was 29.25 ± 1.47 for List 2 (massed practice). The difference between these two means is highly significant, giving a critical ratio of 3.46. Thus it appears that for the same *Ss*, where no selection could enter, distributed practice produced faster learning.

2. The other two tests of the effect of distribution of practice on learning consist of comparing Group 2-D and combined Groups 4-D and 7-D on List 2, and comparing 4-D and 7-D on List 4. In both cases the distributed learning is compared with the massed learning of a group which has been more highly selected and this should reduce the apparent facilitation by distribution. For example, the 4-D and 7-D groups lost more *Ss* than did 2-D. Thus, the selection of *Ss* should result in faster-learning *Ss* being in 4-D and 7-D than in 2-D. For Groups 4-D and 7-D List 2 was learned by massed practice; yet, a mean of 29.25 ± 1.29 trials was required by these groups to learn this list whereas Group 2-D, learning List 2 by distributed practice, required only 20.73 ± 1.29 trials. The difference between these means gives a critical ratio of 4.37.

3. Using the same logic as above, the learning of List 4 by Group 4-D (distributed learning) can be compared with the learning of List 4 by Group 7-D (massed learning). The respective means are 22.04 ± 1.41 and 24.30 ± 1.58. The difference between these means is not significant but its direction favors distributed practice.

In two out of the three comparisons made, distributed practice gave significantly faster learning than massed practice even when the tests made were biased against such results. Thus, while there is no doubt a selection of *Ss* involved, distributed learning appears to contribute to differences in speed of learning above and beyond this selection. To attribute all the difference between massed and distributed learning to selection of *Ss* would require a selection of *Ss* on the first list who were

facilitated by distribution (the "drops" not being facilitated) and whose performance under distributed learning was related to subsequent distributed learning but not to subsequent massed learning. Internal analysis (too detailed to report here) do not support this possibility. It may be concluded that it is highly probable that distributed practice facilitated the learning of these paired-associate consonant lists.

The mean number of trials required to attain successive criteria on the first and last lists is plotted in Fig. 2. As in the serial experiment (**13**), the distributed groups are superior at all points during the learning of the first list but on the last list the massed practice groups are superior early in learning and then change to inferior performance. Figure 2 shows the acquisition of the combined groups on the last list learned. Plotting the acquisition of the last list for each stage of practice shows the same interaction in stage of learning and distribution.

Retention.—Retention of the first and last list was tested 24 hr. after learning. The 168 *Ss* who learned the first list by massed practice recalled a mean of 1.92 ± .073 items while the 168 *Ss* who learned the first list with a 30-sec. intertrial interval recalled a mean of 1.97 ± .076 items. This difference does not approach statistical significance. The mean number of items recalled from the last list after 24 hr. are presented in Table 1. The recall decreases as the number of previous lists increases. The tendency for the distributed group to show better recall is to be expected from the more rapid learning.

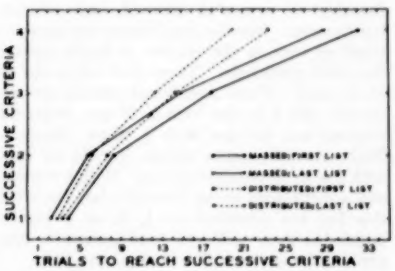

FIG. 2. Trials-to-criteria learning curves for first and last list in the series under massed and distributed practice.

TABLE 1

RETENTION OF THE LAST LIST WHEN LEARNED BY MASSED PRACTICE AND DISTRIBUTED PRACTICE AND PRECEDED BY 1, 3, AND 6 LISTS

Number of Previous Lists	Mean Number Recalled*		Mean Loss Scores**	
	Massed	Dist.	Massed	Dist.
1	1.25	1.46	1.36	1.41
3	.82	1.11	1.74	1.87
6	.75	.73	1.92	2.06

* Estimated σ_M = .125.
** Estimated σ_M = .123.

An analysis which attempts to control for the differences in degree of learning between groups which have learned to the same criterion at different speeds has been presented elsewhere (**3, 12**). From a successive probability analysis of learning a prediction of the number of correct responses for each *S* if he had been given one more trial during acquisition is computed. The observed recall after 24 hr. is subtracted from the predicted score. These loss scores were computed for each group on the first and last lists. The mean loss scores for List 1 of the massed and distributed groups were .88 ± .077 and .97 ± .079, respectively. This difference is not significant (C.R. < 1). The mean loss scores for the last list are presented in Table 1. Again there is a tendency for there

to be less loss in the massed groups, but an analysis of variance of the loss scores gives an F of only 1.01. An F of 12.64 (2 and 330 *df*) indicates that the number of previous lists learned is a significant variable in retention. The greater the number of lists learned the greater the loss over the 24-hr. retention interval. The relearning scores show the same picture as do the recall scores. There is no indication of an interaction between the number of previous lists and the intertrial interval during learning.

Discussion

Hovland (**1**) found that distribution of practice facilitated the learning of paired-associate lists, but other studies (**4, 7, 8**) using a 2:2-sec. rate of presentation have found no effect of distribution on learning paired-associate lists. The material used in the present experiment was consonant syllables of from 50 to 75% association value and the lists had high intralist similarity. With serial lists, where facilitation of learning by distributed practice is the rule rather than the exception, there is no indication that intralist similarity (**6, 9**) interacts with distribution of practice. In the same type of lists distribution is more effective with the low-meaningful material. Therefore it is concluded that the facilitation of learning of paired associates by distributed practice found in this study is due to the low meaningfulness of the consonant syllables.

This study has shown that increased interference from previous learning results in decreased retention, but there is no evidence in the retention data that the intertrial interval during learning interacts with the interference. There were only two levels of distribution, but a supplementary study using a 2-min. interval on the seventh list showed no better recall than after a 4- or 60-sec. interval. It does not seem likely that a longer intertrial interval would change the results. Since there is a decrement in retention related directly to number of previous lists learned comparable to that found in the serial study (**13**), it also seems reasonable to conclude that the lack of interaction at recall between intertrial and number of previous lists was not the result of a lack of interference from previous lists.

The trials-to-criteria curves (Fig. 2) suggest an interaction between intertrial interval and interference during learning. However, the hypothesis that this interaction will also be reflected in the recall scores has received no support from the present study. From the results of the previous serial experiment (**13**) it was suggested that when interlist interference is high, a differentiation of the list being learned from previously learned lists is more effectively accomplished by distributed practice than by massed practice. This differentiation process is suggested by the initially poorer performance under distribution (than under massed) during early learning trials when interlist interference is high. This conception may still be adequate for the rate of loss of such differentiation over time is not known but would be expected to be lost more rapidly with paired-associate lists than with serial lists. A shorter retention interval for the present experiment might give the expected differences in recall.

Summary

This experiment studied the learning and retention of paired consonant syllables as a function of intertrial interval and interlist interference. Interlist interference was manipulated by varying the number of highly similar lists used previously. There were six groups of 56 *S*s each. Three distributed-practice groups learned the first list with a 30-sec. intertrial interval and the last with a 60-sec. interval. Three massed-practice groups learned all lists with a 4-sec. intertrial interval. Both a massed and a distributed group learned a last list after this list was preceded by 1, 3, or 6 highly interfering lists. All lists were learned to one perfect trial with recall occurring after 24 hr.

The results show that distributed practice facilitated learning. This is attributed to the low meaningfulness of the material as compared with material used in previous experiments

where such facilitation was not found. Retention decreased consistently as the number of previous lists learned increased but no difference in retention was observed as a function of intertrial interval during learning.

References

1. Hovland, C. I. Experimental studies in rote learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.
2. Melton, A. W. Materials for use in experimental studies of learning and retention of verbal habits. Mimeo. manuscript, Univer. of Missouri, 1940.
3. Richardson, J., & Underwood, B. J. Comparing retention of verbal lists after different rates of acquisition. *J. gen. Psychol.*, 1957, **56**, 187–192.
4. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
5. Underwood, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, **43**, 329–340.
6. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
7. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.
8. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjective lists as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.
9. Underwood, B. J. Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjective lists. *J. exp. Psychol.*, 1953, **45**, 253–259.
10. Underwood, B. J. Studies of distributed practice: XI. An attempt to solve conflicting facts on retention of serial nonsense lists. *J. exp. Psychol.*, 1953, **45**, 355–359.
11. Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 160–166.
12. Underwood, B. J. Speed of learning and amount retained: a consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.
13. Underwood, B. J., & Richardson, J. Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *J. exp. Psychol.*, 1955, **50**, 39–46.
14. Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received September 4, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

SECONDARY EXTINCTION OF LEVER-PRESSING BEHAVIOR IN THE ALBINO RAT[1]

WILLIAM W. ROZEBOOM[2]

University of Chicago

Interest has recently been revived in the problem of latent, or secondary extinction,[3] a problem which may be formulated: If an instrumental response (R) has been conditioned to a stimulus (S_o) by following the occurrence of R in the presence of S_o with a reinforcing stimulus (S^r), and if the reinforcement value of S^r is subsequently modified, will the response to S_o remain unaffected or will it have altered in accordance with the new reinforcement value of S^r? Existence of secondary extinction would seem to imply that instrumental behavior involves a mediation process whose behavioral effects are governed by those of the reinforcer through which the behavior is established (**11**).

Other than Miller's (**9**) demonstration that shocking rats placed directly in a distinctive goal box tended to suppress a running response which had previously terminated in that goal box, evidence either for or against secondary extinction is surprisingly meager. Seward and Levy (**12**) and Deese (**4**) have reported positive findings, but both experiments are complicated by a differential secondary reinforcement of experimental and control groups (**2, 10**), an effect whose importance has clearly been shown by Moltz (**10**), who was unable to demonstrate positive results in its absence, and also more recently by Coate (**3**). Negative findings have been reported by Bugelski, Coyer, and Rogers (**2**) and by Scharlock (**13**), but since these experiments also failed to show a secondary reinforcement differential which should have appeared, their significance is problematic.

The possibility of secondary extinction is here explored in thirst-motivated lever pressing of the albino rat. Such behavior, aperiodically reinforced by water reward, may somewhat arbitrarily be analyzed as a response chain of three main links: (*a*) lever pressing, (*b*) approach to the water cup, and (*c*) drinking. Stimuli accompanying water delivery ("magazine stimuli") serve to reinforce lever responses and provide discriminative cues for approach to the water cup (see **5**). The question then arises whether, after the lever habit is established, altering the secondary reinforcement value of the magazine stimuli affects the strength of the lever habit prior to culmination of the response in the modified secondary reinforcement. The reinforcement properties of magazine stimuli may be altered, while levers are absent, in at least two ways ("magazine reconditioning"): (*a*) extinction of positive value through repeated activations of

[1] The present data are taken from a dissertation submitted to the Faculty of the Division of the Biological Sciences, the University of Chicago, in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his appreciation for the valuable assistance of his sponsor, Dr. Howard F. Hunt, and for the generous support of the National Science Foundation, whose fellowship program has subsidized the writer's graduate studies. This research was also supported in part by the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, under Contract No. DA-49-007-MD-291.

[2] Now at the University of Minnesota.

[3] The title, "nonresponse-extinction" has been proposed by Deese (**4**) for this design. However, this is also appropriate to extinction procedures (e.g., drug-induced response blockage) considerably different from the present paradigm. Nor is the term "latent extinction" quite appropriate here, since there is no period during which the extinction is latent in the usual sense of the word. The term "secondary extinction" is here offered in analogy to "secondary generalization," a phenomenon of similar theoretical import (see **11**).

the dry water magazines (Exp. I); or (*b*) endowment with secondary aversive properties by pairings of magazine stimuli with shock (Exp. II). The extent, if any, to which magazine reconditioning has weakened lever responding may then be determined by replacing the levers and comparing experimental and control groups for unreinforced lever presses ("lever extinction"). Whether magazine reconditioning was, in fact, accomplished may subsequently be tested by reconnecting the dry water magazines to the levers. Comparison of additional lever presses ("magazine extinction") reveals the residual reinforcement properties of the magazine stimuli.

Experiment I

Method

Subjects.—The *S*s were male albino rats of the Sprague-Dawley strain. All were about 100 days old and experimentally naive at the beginning of the experiment. Forty *S*s were given initial training, from which two matched groups of 16 *S*s each were selected at the beginning of magazine reconditioning.

Maintenance and running schedule.—The *S*s were run daily for a 20-min. recorded interval, plus about 1½ min. both before and after the interval recorded. They were always run under a 23-hr. water deprivation, water being received once a day for a period of 25 min. beginning 25 min. after return to the home cage. Dry food pellets were always available in the home cage.

Apparatus.—The Skinner boxes were a battery of eight in regular use at the University of Chicago. These boxes are sound and light insulated, with electrifiable grill floors. The water delivery mechanism consists of a small cup which dips into an underlying reservoir when activated. A single horizontal T lever is located above and 3 in. to the right of the water cup. Throughout the present experiment, reinforcement was controlled through a timing device which makes reinforcement available on a prearranged semi-random interval schedule. For the rates at which lever pressing stabilized, the schedule used furnished an average of one reinforcement for seven lever presses. Activation of the water magazine produced two prominent clicks, over-all duration of which was approximately .5 sec. In addition, a small, dim, neon bulb to the left of the water cup flashed on for .1 sec. at the beginning of magazine activation.

Pretraining.—For the first six days of training, *S*s were run with disconnected levers to familiarize them with the apparatus and provide a record of unconditioned operant lever-pressing level. The latter variable showed little or no relation to subsequent behavior measures.

Conditioning.—Three types of training were given during conditioning. (*a*) On 20 "lever conditioning" days (L), levers were connected to the full water magazines, with reinforcement available on the schedule reported. (*b*) On each of eight "magazine conditioning" days (M), levers were removed and magazines were activated by *E*, 20 activations per *S* on each of the first four days and 10 per *S* on each of the last four. The purpose of M was to establish and maintain the secondary reinforcement value of magazine stimuli independently of lever presence. (*c*) On five "adaptation" days (A), levers were removed and no reinforcements were given. This was to sharpen discrimination of lever stimuli as the essential S^D's for lever pressing. The sequential order of training was: 4M, 4L, M, 4L, M, 2L, 5A, 2L, M, 4L, M, 4L. An average response rate of around 200 lever responses per run was achieved by the fifth L day, gradually increasing to a daily mean of 250 presses, averaged over the last four days. There was little inter-*S* variation in total number of reinforcements, and later statistical analysis revealed no correlation between this variable and other behavior measures.

Magazine reconditioning.—After Day 20 of lever conditioning, an experimental and a control group of 16 *S*s each were selected, closely matched for terminal response output, as measured by the last four days of conditioning. With levers removed, the experimental group then received a total of 400 activations of the dry magazine, distributed over 25 runs in the sequence: 40, 40, 0, 40, 40, 0, 0, 20, 20, six days not run, 40, 20, 20, 0, 20, 20, 0, 0, 20, 20, 0, 0, 0, 0, 20, 20. Spacing of extinction was designed to encourage spontaneous recovery and subsequent re-extinction. Control *S*s were treated during magazine extinction exactly as were the experimentals except that controls received no magazine activations and hence no extinction of the positive secondary reinforcement value of magazine stimuli.

Lever extinction.—Both groups were then run for 16 days with levers replaced but disconnected from the water magazines. By the end of this time, response rates had returned to their original unconditioned operant level.

Magazine extinction.—All *S*s received a terminal eight runs wherein levers were recon-

nected to the dry magazines, with magazine stimuli contingent upon lever pressing according to the same schedule encountered during conditioning.

Results and Discussion

Examination of Fig. 1 reveals the course of response output during lever extinction to be practically identical for both groups, with the experimentals, contrary to hypothesis, actually showing a greater total output than controls. (The unusually large difference on Day 14 was primarily the contribution of a single *S*, who responded in a highly atypical manner that day.) The only evidence remotely favoring the existence of a stronger response tendency in the control group is its slightly greater output for the first two days. To test the significance of this, the sum of the first two days' output was subjected to an analysis of covariance (**14,** p. 378 ff.), even though *post hoc* selection of variables in this way biases statistical tests toward spurious significance. When the regression of the first two days' lever extinction output on terminal

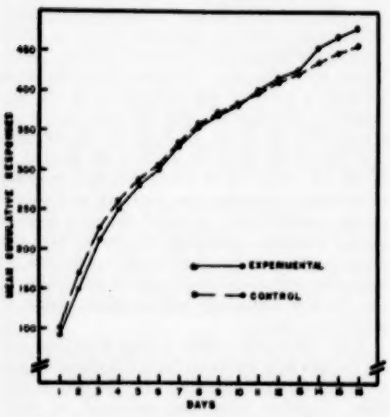

FIG. 1. Cumulative mean lever responses during "lever extinction" in Exp. I.

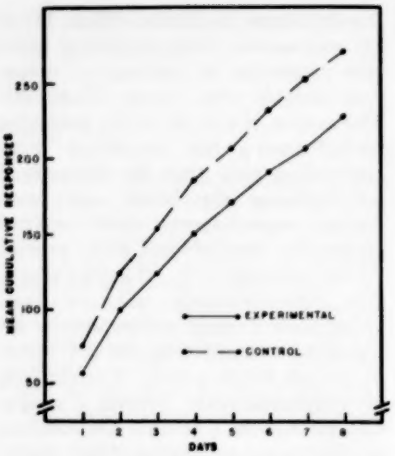

FIG. 2. Cumulative mean lever responses during "magazine extinction" in Exp. I.

conditioning rate is accounted for ($r = .54$), the ratio of between-group to within-group variance is $F = .96$, which occurs under the null-hypothesis with probability $P = .36$ (1 and 29 *df*). During magazine extinction (Fig. 2), however, a consistent difference appears. Analysis of covariance applied to the first day's output gives $F = 7.00$ ($P = .01$, 1 and 29 *df*). Thus, while there is no detectable difference between experimental and control groups in lever extinction outputs, reintroduction of magazine stimuli contingent upon lever pressing reveals clearly a greater residual reinforcement value of magazine stimuli for control *S*s.

It would appear that under the parameters of this experiment, at least, partial extinction of a secondary reinforcer with whose aid an instrumental response has been established has no effect on the strength of the habit. It might be protested that the secondary reinforcers were insufficiently extinguished for the experimental group to display a detectably weakened habit, for magazine

TABLE 1

MEASURES OF PERFORMANCE, EXPERIMENTS I AND II

Group	*N*	Terminal Cond. Rate		First 2 Days Lever Extinction			Total Lever Extinction			First Day Magazine Extinction		
		Mean	*SD*	Mean*	*SD*	Adj. *SD*†	Mean*	*SD*	Adj. *SD*†	Mean*	*SD*	Adj. *SD*†
Experiment I												
E	16	241.2	63.1	150.6	53.6	50.2				57.0	22.2	22.6
C	16	240.7	62.9	169.1	68.4	58.8				75.9	19.1	18.6
Experiment II												
E	22	242.8	80.7	107.8	44.1	42.3	236.9	87.8	84.2	37.6	35.3	—
C_1	16	245.4	76.0	121.4	36.6	32.7	272.4	64.8	61.3	93.9	31.2	25.2
C_2	16	244.8	82.7	108.5	43.2	34.0	231.1	101.8	83.6	73.5	25.2	22.3

* Since groups were matched for terminal conditioning rate, adjusted means differ only insignificantly from observed means.

† Best estimate of parametric *SD* after removal of variance component attributable to terminal conditioning rate.

reconditioning produced a reinforcement loss of only about 25% (based on total responses during magazine extinction less an operant level base line of 10 responses per day), even though Beckwith's data (1) suggest that less than half this number of dry magazine activations following lever responses would have been ample for complete extinction. The fact remains, however, that the reinforcement difference, though small, was still large enough to be clearly evident statistically. There is no apparent reflection of this difference during lever extinction, and it must be concluded that the strength of a lever habit is at best quite insensitive to changes in reinforcement value of the stimuli by which the habit was initially reinforced, at least under the present parameters of conditioning.

EXPERIMENT II

Experiment II was conducted in the hope that the behavioral effect of a secondary reinforcer could be altered more radically by pairings with a primary aversive stimulus, and hence more likely reveal secondary extinction if this phenomenon exists.

Method

Subjects.—The *S*s were similar to those of Exp. I, a total of 54 being used. No *S*s were lost or discarded during the experiment.

Apparatus, maintenance, and running schedule.—Identical to Exp. I.

Pretraining and conditioning.—Treatment of *S*s in Exp. II for these phases was, with three minor exceptions, identical to that of Exp. I: (*a*) *S*s were given only four pretraining runs; (*b*) only three days of magazine conditioning were given prior to the first day of lever conditioning; and (*c*) to facilitate later discriminations, on each of the last four magazine conditioning days, all *S*s also received 10 presentations of the tone later experienced by Group C_1 during magazine reconditioning. The mean daily response output over the last four days (terminal conditioning rate) was 244, very similar to the corresponding rate in Exp. I.

Magazine reconditioning.—The *S*s were then divided into three groups closely matched for terminal conditioning rate; an experimental group (E) of 22 rats, and two control groups (C_1, C_2) of 16 rats each. As in Exp. I, levers were absent throughout magazine reconditioning, and the water magazines were always dry.

During the first four days, each *S* in Group E received 120 presentations of the dry water magazine, 40 on each of the first two days and 20 on each of the others. Groups C_1 and C_2 were run as usual, but received no magazine stimuli. During the next 22 runs, each *S* of

every group received a total of 40 electric shocks. The number of shocks varied from day to day according to the sequence: 3, 2, 4, five days not run, 3, 1, 2, 4, 0, 0, 4, 2, 0, 3, 0, 1, 2, 1, 0, 0, 4, 4, 0. The shock circuit was set to deliver an alternating current of approximately 1.0 ma. for .2 sec., automatically timed. For Group E, each shock was preceded by activation of the dry water magazine, with an automatically timed interval of .5 sec. between cessation of magazine stimuli and onset of shock. Group C_1 had each shock preceded by a medium intensity, 1000-cps tone, of .5 sec. duration and with an interval of .5 sec. between tone offset and shock onset. Group C_2 received shock alone, unheralded by any warning signal.

Lever and magazine extinction.—As in Exp. I, all *S*s were now run with levers replaced, but disconnected. When lever extinction outputs appeared asymptotically equal for all groups, the dry magazines were reintroduced for seven days of magazine extinction.

Results and Discussion

Although the response outputs of Groups E and C_2 during lever extinction can scarcely be distinguished in Fig. 3, the record of C_1 shows a distinct superiority. To test the statistical significance of this difference, two preselected measures—sum of responses for the first two days and total lever extinction presses—were separately subjected to an analysis of covariance in order to control for the variance component attributable to terminal conditioning rate, upon which the groups were matched. The between-group to within-group variance ratio for Groups E and C_1 in total lever extinction output is $F = 1.98$, which, for 1 and 35 *df*, gives $P = .20$ that this great a mean difference could have occurred by chance. For the total lever extinction output of Groups C_1 and C_2, $F = 2.46$, for which $P = .14$ ($df = 1$ and 29). The corresponding chance probabilities of the differences in the first two days of lever extinction are somewhat higher ($P = .32$ and $P = .30$, respectively). The data thus fall short of conventional statistical significance and do not warrant the conclusion that the theoretical populations, of which E, C_1, and C_2 are samples, differ in lever extinction output. However, neither do these data *support* the conclusion that there are no genuine group differences, for failure to reject the null-hypothesis is not in itself grounds for its acceptance, and the observed differences are too large to be completely ignored. Under the null-hypothesis, a difference between C_1 and E as large as the one observed and in this direction, as predicted, would be expected only once in ten, which, in the absence of further evidence, is a precarious statistic upon which to rest the conclusion that Groups E and C_1 did, in fact, show equal parametric response tendencies during lever extinction. Of much less ambiguous interpretation are the differences during magazine extinction (Fig. 4), as revealed by analysis of the first day's outputs. Covariance analysis yields a between-group to

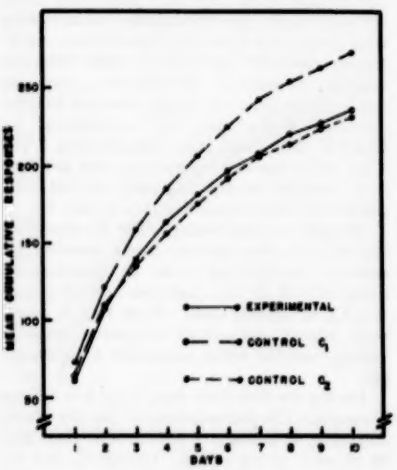

FIG. 3. Cumulative mean lever responses during "lever extinction in Exp. II.

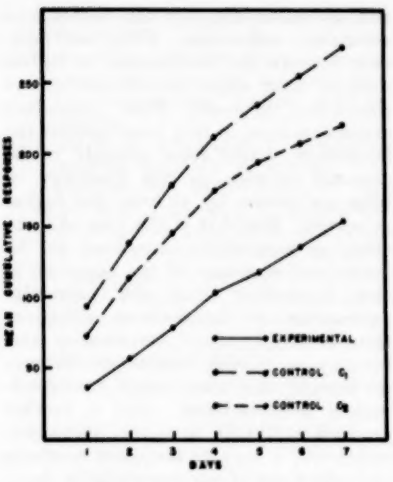

FIG. 4. Cumulative mean lever responses during "magazine extinction" in Exp. II.

within-group variance ratio of $F = 5.65$ for Groups C_1 and C_2, for which $P = .02$ ($df = 1$ and 29), while a χ^2 comparison of scores dichotomized at the common median reveals the inferiority of E to be highly significant ($P < .001$).

Before these results can be evaluated, it is necessary to judge the extent to which comparison of lever extinction outputs is here a fair comparison of response tendencies. As in Exp. I, the only external stimuli present during lever extinction with contingency histories differing among the groups were the general environmental stimuli. In contrast to Exp. I, however, there is substantial empirical evidence from which here to anticipate differences in lever extinction outputs due to environmental stimuli alone, irrespective of any additional secondary extinction effects. It is a solidly established fact (e.g., **7**) that stimuli associated with aversive stimuli in the Skinner box subsequently suppress lever pressing. (A plausible explanation is that such stimuli, having become secondarily aversive through type-S conditioning, now elicit "anxiety" responses —crouching, defecating, etc.—which compete with other behavior.) This general suppressant effect may clearly be seen in the present data by comparison of the lever extinction outputs of the control groups in Exps. I and II.

Merely giving experimental and control *S*s an equivalent amount of shock, however, is insufficient to equate the behavioral effects of environmental stimuli, since experimental *S*s also received magazine stimuli as a signal for shock. It is to be expected both on theoretical (**6**) and experimental (**8**) grounds that a warning signal prior to an aversive stimulus should decrease aversive conditioning of the environmental stimuli, since the latter then form only a part of the total secondary aversive compound. This effect may be called, in a purely descriptive sense, the "buffer" effect. For present purposes, the buffer effect implies that magazine stimuli preceding shock should decrease the subsequent suppression of lever responding by environmental stimuli. For this reason, two control groups were run, one (C_1) receiving a warning tone before shock, and the other (C_2) receiving none. It was felt that Group C_1 should be most comparable to Group E for the behavioral influences of environmental stimuli, while Group C_2 would check the importance of the warning signal. While the difference between C_1 and C_2 in Fig. 3 does not quite attain statistical significance, Fig. 4 reveals that with high likelihood, a warning signal before shock does tend to lessen the aversive conditioning of environmental stimuli.

The possibility must also be considered, however, that tone might have been more effective than magazine stimuli as a buffer, and hence have produced a lever extinction superiority of Group C_1 over Group E apart from any secondary extinction effects. There are, unfortunately, no experimental data on the determinants of buffer efficiency. For present purposes, an attempt was made to evaluate relative "anxiety"

during lever extinction by consideration of defecation rates. The somewhat involved analysis (11) points to an "anxiety" ordering of $C_1 < E < C_2$, with Group C_2 the most disturbed as predicted, but with Group E as close or closer to Group C_2 than to Group C_1. Presuming an approximately linear relationship between degree of "anxiety" and response suppression over the range concerned, this finding suggests that at least half the observed difference between Groups C_1 and E, already of questionable significance, may be due to emotional differences. It must therefore be concluded that, although the possibility cannot entirely be ruled out that pairing magazine stimuli with shock might have produced a slight loss of lever-pressing strength, the effect is scarcely detectable, at best of only minor importance in the composition of the habit.

Failure to detect any appreciable effect, moreover, can scarcely be attributed to inadequate reconditioning of magazine stimuli. The differences between Group E and Group C_1 or Group C_2 during magazine extinction are even greater than shown in Fig. 4. The distribution of Group E was markedly skewed, with but three *S*s (14%) accounting for 42% of the group's total output on the first day. Moreover, a typical *S* in Group E showed a positively accelerated daily cumulative magazine extinction curve, clearly indicating extinction of the secondary aversive properties of the magazine stimuli and subsequent reinstatement of their yet unextinguished positive reinforcement value. In contrast, *S*s of both control groups gave conventional negatively accelerated extinction curves. Obviously, then, pairings with shock were highly successful in endowing magazine stimuli with aversive properties.

Experiments I and II are in complete agreement that under at least some parameters of learning, an instrumental response, once conditioned, is little if at all affected by subsequent modifications in value of the reinforcers by which it was established. The present data do not, of course, disprove the existence of secondary extinction. They do, however, indicate the phenomenon to be less evident than might be anticipated on theoretical grounds. That secondary extinction does, in fact, exist outside the laboratory would seem scarcely to be doubted in view of the flexibility of behavior shown by at least the higher primates. But if it is the case that an effect so potentially important in the behavioral economy of the organism is this dependent upon the particular parameters of the learning situation, then investigation of the identity and functioning of such parameters becomes of considerable importance to experimental behavioristics. And if further research still fails to reveal secondary extinction, it may be necessary to revise radically some of our expectancies about the constitution of behavior.

Summary

An attempt was made to discover secondary extinction in thirst-motivated lever pressing of the albino rat. A water-reinforced lever habit was established in thirsty rats, with stimuli accompanying water delivery (magazine stimuli) mediating reinforcement of the habit. The effect on the lever habit of modification in the reinforcement value of magazine stimuli was then determined. Two experiments are reported, one in which simple extinction of magazine stimuli as positive reinforcers was attempted, and another in which magazine stimuli were also aversively conditioned. Neither experiment resulted in any apparent degree of secondary extinction.

References

1. Beckwith, W. C. A study of some factors affecting extinction after partial reinforcement. Unpublished doctoral dissertation, Univer. of Chicago, 1956.
2. Bugelski, B. R., Coyer, R. A., & Rogers, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.*, 1952, **44**, 27–30.
3. Coate, W. B. Weakening of conditioned bar-pressing by prior extinction of its subsequent discriminated operant. *J. comp. physiol. Psychol.*, 1956, **49**, 135–138.
4. Deese, J. The extinction of a response without performance of the choice re-

sponse. *J. comp. physiol. Psychol.*, 1951, **44,** 362–366.

5. Estes, W. K. A study of motivation conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1949, **39,** 306–310.
6. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60,** 276–286.
7. Hunt, H. F., & Brady, J. V. Some effects of electro-convulsive shock on a conditioned emotional reaction ("anxiety"). *J. comp. physiol. Psychol.*, 1951, **44,** 88–98.
8. Hunt, H. F., & Brady, J. V. Some effects of punishment and intercurrent "anxiety" on a simple operant. *J. comp. physiol. Psychol.*, 1955, **48,** 305–310.
9. Miller, N. E. A reply to "Sign-gestalt or conditioned reflex?" *Psychol. Rev.*, 1935, **42,** 280–292.
10. Moltz, H. Latent extinction and the reduction of secondary reward value. *J. exp. Psychol.*, 1955, **49,** 395–400.
11. Rozeboom, W. W. Mediation processes in learning. Unpublished doctoral dissertation, Univer. of Chicago, 1956.
12. Seward, J. P., & Levy, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39,** 660–668.
13. Scharlock, D. P. The effects of a pre-extinction procedure on the extinction of place and response performance in a T maze. *J. exp. Psychol.*, 1954, **48,** 31–36.
14. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

(Received September 4, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

ERROR REINFORCEMENT IN A MODIFIED SERIAL PERCEPTUAL-MOTOR TASK [1]

MELVIN H. MARX AND ROBERT A. GOLDBECK [2]

University of Missouri

A recent study by Marx (**2**) reported evidence supporting a reinforcement explanation of the spread-of-effect phenomenon. A serial perceptual-motor task was used in an attempt to eliminate the usual response-bias tendencies characteristic of the number-guessing experiments. This task also had the advantage of permitting a more sensitive measure of response strength in that measurement of error gradients was not dependent upon exact repetition of errors (cf. **1**).

The purpose of the present study was to confirm the critical role of reinforcement in the production of after-error response strength, using a modified version of the task earlier used by Marx. The modification was intended to eliminate response biases associated with the central rest position of the apparatus.

Method

Subjects

The *S*s were 379 male volunteers from the freshman and sophomore Air ROTC classes at the University of Missouri.

Apparatus

The apparatus was a modified Mashburn Complex Coordinator (**5**). A wood panel with a white paper covering was mounted on the front of the regular Mashburn panel. This special panel contained two rows of 13 lights which showed through the paper covering. The upper row of lights was seen through a green filter. Each vertical pair of lights was framed by wood strips, forming 13 boxes. The *S* could light any one of the 13 white lights by the left-right movement of the Mashburn control stick. The *S* could turn on the green light in the box above any selected white light by pushing the Mashburn rudder bar to the extreme right position. By means of a switch hidden from *S*, *E* could either allow the green light to come on or a buzzer to sound when *S* pushed the rudder bar.

There was no spring tension on the stick for left-right movement; the usual central resting position was thus eliminated.

Procedure

The task was represented to *S* as a position memory test; he was told that the object of the test was to remember correct responses which could be discovered only by chance. The *E* called out the numbers 1 through 15, which served as the stimulus items. When *E* called out a stimulus number, *S* moved the stick to light one of the white lights.

Phase 1.—The *S* moved the stick from the extreme right position to make each response. Holding the stick in position, *S* then pushed the rudder bar to determine whether the light was the correct one for that stimulus number. A correct response was signaled by the flashing of the green light; the sounding of the buzzer designated an incorrect response. The *E* called out the next stimulus number approximately 1 sec. after *S* returned the stick to its position at the right.

A trial consisted of the responses to the 15 numbers. Trials were separated by *E*'s statement that the 15 numbers would be gone through again and that any correct responses were to be repeated.

The *S*s were discouraged from using any system in choosing lights. As an aid for re-

[1] We would like to acknowledge the cooperation of the Air ROTC unit at the University of Missouri in providing the *S*s for this experiment. This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25631, monitored by the Perceptual and Motor Skills Research Laboratory, Human Resources Research Center. Permission is granted for reproduction, translation, publication, use, and disposition in whole and in part, by or for the United States Government.

[2] Now at American Institute for Research, Pittsburgh, Pennsylvania.

membering light positions without counting their distance from the ends, paper strips were placed beneath the four lights on each end of the row.

The experimental and control groups each consisted of 150 *Ss* who provided usable data for this phase. The *Ss* in the control group were given two trials without reward; each response was followed by the buzzer. If none of the responses to numbers 1 through 12 was repeated from Trial 1 to Trial 2, a third trial was given. If *S* produced more than one repetition, the one occurring earliest in the list of numbers was used. A total of 22 *Ss* failed to repeat a single response under these conditions and so could not be used in this phase of the experiment.

For each control *S*, a paired *S* in the experimental group was given a reward following his response to the same stimulus number. In order to make the groups as comparable as possible only *Ss* who repeated this rewarded response were used as experimental *Ss* in this phase of the experiment. A total of 57 *Ss* had to be replaced for failure to repeat the correct response before the necessary 150 experimental *Ss* were obtained.

Phase 2.—Following the trials given in Phase 1, all *Ss* were given two additional trials. For these trials, *Ss* were instructed to move directly from one light to the next without moving the stick back to the right after each response, as they had done on the earlier trials.

Under these conditions, 61 *Ss* from the control group in Phase 1, who had received no reward, were rewarded for the response to Stimulus 14 or 15 on the two trials of Phase 2. The reward was given to maintain motivation and morale. This procedure allowed for the occurrence of chance repetitions within the first 10 stimulus numbers which would not be influenced by the aftereffects of reward and so could be used as controls.

All other *Ss* were given two rewards on each trial. The *Ss* from the experimental group of Phase 1 were rewarded for the response to the same stimulus number for which reward had earlier been given. This was done to insure continuity of reward for *Ss* who repeated previously rewarded responses. The 57 *Ss* eliminated from the Phase 1 analysis because of failure to repeat the rewarded response were treated in the same manner as the other experimental *Ss*. The *Ss* from the control group of Phase 1 were rewarded for the response to the stimulus number for which the critical chance repetition had earlier occurred. This was done to equate reinforcement conditions with the *Ss* from the Phase 1 experimental group. In all cases the second reward was given for a response to a stimulus number far enough removed from the first reward to be free of its aftereffect.

Results

The basic measure of response strength used to evaluate error strength was the difference, in terms of light units, between the responses made immediately following the rewarded response and its repetition on the next trial. This difference score for the first after-error ranged between 0 and 12, with a low difference score indicative of high response strength.

Phase 1

The mean difference score for the first after-error was 2.79 for the experimental group, and 3.53 for the control group. A matched-pairs *t* test showed this difference in favor of the experimental group to be significant ($t = 2.42$, $P < .02$ for 148 *df*). This finding indicates that reward increased the tendency to repeat the following response over and above any increment associated with preexperimental response sequences.

It was considered possible that the role of reward was to produce some special response bias on both trials. The correlation between first- and second-trial responses was therefore obtained for the first after-error. Because a group response bias would not produce intertrial ordering of scores, the correlation coefficient should not be sensitive to the restricted range of responses produced by such a response bias. The obtained *r*'s for the experimental and control groups respectively were .113 and $-.024$, with .159 needed for significance at the .05 level. The *t* for the difference between these two correlations was 1.17, with 1.96 needed at the .05 level. Failure to

obtain evidence of a significant correlation for the experimental group leaves open the possibility that the tendency of this group to repeat the first after-error response might be attributed to a response bias produced by reward.

Phase 2

In order to use only completely new data, difference scores were computed only from the 164 experimental *Ss* who repeated the rewarded response to stimulus numbers not used for reinforcement in Phase 1.

For the 61 control *Ss*, receiving only one reward at the end of the list, data following the first repeated response were used, but only if repetition occurred at least four stimulus steps before the rewarded response. The 43 *Ss* meeting this criterion constituted the control group.

The mean difference score for the first after-error was 2.54 for the experimental group, 4.21 for the control group. This difference was clearly significant ($F = 13.75, P < .01$ for 1 and 205 *df*).

To determine if the low difference score for the experimental group could be attributed to a response bias associated with reward, the correlation between intertrial first after-error responses was obtained. An r of .291 ($P < .01$) lends support to an individual-strengthening interpretation of the data. The obtained r for the control group was $-.235$, short of significance (.304 needed at .05 level). The t for the difference between these two r's was 3.05 ($P < .01$).

To rule out the possibility that the correlation between responses following reward was due to some correlation between the rewarded response and the response following it on both trials, these correlations were partialled out of the after-error correlation for the experimental group. The partial r of .276 obtained was still significant at the .01 level.

Discussion

The present results, particularly those in Phase 2, offer further support for the reinforcement interpretation of the error-reinforcement, or spread-of-effect, data. As in previous research reporting positive results (**2, 3, 4**), the use of repeated nonrewarded responses as controls and application of the partial-correlation technique have served to eliminate the effect of intratrial influences such as are emphasized in the usual guessing-sequence type of artifactual explanation (cf. **1** for a critical review of this problem). The present results may be more particularly offered in support of the modified Thorndikian interpretation (**2**), which stresses the role of the repeated rewarded response as a stimulus factor as well as a reinforcer for the first after-error.

In the present experiment only the first after-error difference scores were analyzed. This was done on the ground that the first after-error offers the best and simplest opportunity for positive evidence; such evidence, if adequately controlled, is all that is needed to establish the reinforcement interpretation. The more complex question of the complete gradient effect, with which Thorndike was originally concerned, can better be handled later.

The implications of the more consistent statistical evidence found in Phase 2 are not clear. The sequential resting-position procedure and the use of later trials were confounded. It is not possible to say on the basis of the present data which, if either, of these factors was responsible for the more clear-cut positive results.

The fact that fewer control than experimental *Ss* were eliminated from the analysis in each phase reflects the

necessary difference in treatment (control *Ss* merely required to have a chance intertrial repetition, experimental *Ss* required to repeat a correct response). It was necessary to have the critical response repeated in the experimental group so that the response-bias factor would be present in each case and any difference could then be attributed to the addition of the reinforcement factor. If there was any selection bias introduced from differential sampling it would appear to have been against the reinforcement hypothesis. Elimination of experimental *Ss* who failed to repeat a correct response would tend to eliminate the less strongly motivated and less capable retainers. The remaining *Ss*, as better retainers, would also be expected to recall and vary more of the errors made and therefore be less likely to show the spread effect. It is, of course, impossible to say on the basis of the present data to what extent a selection bias was operating, but the important point is that if present it would seem to work against, rather than for, the hypothesis.

Summary

An experiment was designed to control for response biases produced by the central rest position of an apparatus used in an earlier study which had given positive spread-of-effect results. Two methods of correction were used: (*a*) the extreme right position was used as the rest position; (*b*) the previous response position served as the rest position. The task was presented to 379 *Ss* as a position memory test. Typical prearranged spread-of-effect reinforcement conditions were employed. Response-bias factors were controlled by the use of error scores following repeated nonrewarded responses and by the application of the partial-correlation technique.

Errors made immediately following repeated rewarded responses were significantly stronger than errors made immediately following repeated nonrewarded responses. These results were seen as offering further support for the reinforcement interpretation of spread-of-effect data.

References

1. Marx, M. H. Spread of effect: a critical review. *Genet. Psychol. Monogr.*, 1956, **53**, 119–186.
2. Marx, M. H. Gradients of error reinforcement in a serial perceptual-motor task. *Psychol. Monogr.*, 1957, **71**, No. 8 (Whole No. 437).
3. Marx, M. H. Gradients of error-reinforcement in normal multiple-choice learning situations. *J. exp. Psychol.*, 1957, **54**, 225–228.
4. Marx, M. H., & Goodson, F. E. Further gradients of error reinforcement following repeated reinforced responses. *J. exp. Psychol.*, 1956, **51**, 421–428.
5. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Res. Rep. No. 4. Washington: U. S. Govt. Print. Off., 1947.

(Received October 1, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

CONDITIONING OF MUSCLE ACTION POTENTIAL RESPONSES RESULTING FROM PASSIVE HAND MOVEMENT[1]

DONALD G. DOEHRING[2]

Indiana University

Passive movement can be defined as displacement of part of an organism produced by the action of an agent other than the organism itself. As commonly used, the term seems to imply that the organism does not actively participate in the movement.

The conditioning of passive movement has been attempted in two experiments. Miller and Konorski (**6**) found that when a stimulus was paired with the passive lifting of a dog's foot, and this combination was followed by the presentation of food, the stimulus came to serve as a CS for an active flexion response. Woodbury (**9**) used a similar procedure, modified to prevent the possibility of conditioned avoidance movement or postural adjustment, and failed to obtain a CR.

Conditioning should take place if a stimulus is paired with a response and the response is followed by a reinforcing state of affairs, according to most learning theories. From Woodbury's results, then, it might be concluded that passive movement does not involve response by an organism. Passive movement would thus be qualitatively different from reflex or voluntary movement.

If passive movement does involve organismic responses, however, these responses would probably be relatively small. Any CR related to these responses would probably be even smaller. The measure of response (treadle movement) used by Woodbury may not have been sensitive enough to detect such small responses. The recording of action potentials from muscles directly affected by passive movement should provide a more sensitive measure in an investigation of the conditioning of passive movement.

In the present experiment passive hand movement was paired with a CS, and muscle action potentials were recorded from muscles most likely to be involved in the passive movement. The problems to be investigated were: (*a*) Do action potential responses occur in muscles directly affected by passive hand movements? (*b*) Can conditioned muscle action potential responses be developed in situations where the UCS is a passive hand movement?

Method

Experimental design.—The conditioning procedure was patterned after the Pavlovian paradigm. The CS was a weak 2-sec. white noise for all groups, whose termination was immediately followed by the UCS on conditioning trials. Two experimental groups of 10 *Ss* each were used, with one group (XA) receiving passive extension of the right hand and the other group (XB) receiving passive flexion of the right hand as the UCS. Corresponding control groups (CA and CB) of 10 *Ss* each received auditory stimuli attendant upon the production of passive movement without receiving the actual movement. Trials 1, 8, 12, 15, and 21 through 23 were test trials, where the CS was presented alone. Trials 2–7, 9–11, 13,

[1] This article is based upon a dissertation submitted to the department of psychology at Indiana University in partial fulfillment of the requirements for the Ph.D. degree. Thanks are due to Dr. R. C. Davis who directed the dissertation.

[2] Now employed by Central Institute for the Deaf, St. Louis, Mo.

14, and 16–20 were conditioning trials, where the CS was paired with the UCS.

Subjects.—The *Ss* were 40 students from introductory courses in psychology at Indiana University, none of whom had served in previous muscle action potential experiments.

Apparatus.—The equipment and techniques used for the measurement and recording of muscle action potentials have been previously described by Davis (**1, 2, 3**). Briefly, muscle action potentials are picked up by recording electrodes, amplified, integrated over .1-sec. intervals, further amplified, and recorded by mechanical recording pens. The height of each recorded peak represents the total microvolts of potential within a given .1-sec. interval. One change made in the usual recording procedure was the use of bipolar electrodes in place of monopolar electrodes in order to obtain a more precise measure of the difference between extensor and flexor levels of response. Recordings from each muscle were made by the use of a pair of ¼-in. brass electrodes mounted ½ in. apart on a piece of Plexiglas. Since the distribution of the logarithms of microvolts of action potential has been found to be more nearly normal than that of unconverted measures (**5**), all action potentials were measured on a 100-log microvolt scale.

Apparatus for producing passive movement included a 21-in. × 4-in. strip of wood covered with sponge rubber, on which *S* rested his right forearm. The strip was fastened to the top of a table in the experimental room. The front section of the strip was movable, being hinged at a point directly under *S*'s wrist. Two coil springs were fastened to the front of the movable section, each exerting a force of 29 oz. when the section was in a horizontal position. The movable section was held down by a metal rod which was connected at the other end to a pair of solenoid plungers, the coils of which were mounted on the outside of the experimental room. When the solenoids were activated, the bar was withdrawn from a roller bearing mounted on the underside of the movable section, allowing it to be raised by the springs to an angle of about 30° from the horizontal. The movable section could then be returned to a horizontal position by pulling a string from the outside of the experimental room. Reinsertion of the bar maintained this position until the solenoids were again activated.

The CS was a 37 db white noise produced by a white noise generator and delivered through a loudspeaker mounted in a wall of the experimental room. The white noise could be heard easily by all *Ss*, but was not loud enough to produce a large muscle potential response by itself. The duration and temporal contiguity of the white noise and solenoid activation were controlled by an electronic timer (**8**). A continuous 35 db low-frequency noise was provided by the hum of a loudspeaker located directly below the speaker which delivered the white noise. This served to mask extraneous sounds from outside the experimental room.

Procedure.—The *Ss* were divided randomly into four groups of 10 each. Their order of running was randomized.

Muscle potentials were recorded from two muscles on the right forearm, the extensor digitorium and flexor digitorium. The muscles were located by palpation, and a pair of electrodes was fastened over the crest of each muscle after electrode paste had been applied.

The *S* was seated in a darkened, sound-insulated experimental room with his right arm loosely fastened to the sponge rubber covered strip of wood in such a way that his wrist was directly over the hinge of the movable section. For Groups XA and CA the palm of the hand was down, and for Groups XB and CB the palm was up. The *S* was told that the purpose of the experiment was to investigate the effects of several types of distraction and was asked to remain as relaxed as possible. The first trial was given about 10 min. after *S* had been placed in the experimental room, and the intertrial interval was varied unsystematically between 40 and 60 sec.

The CS was presented to all groups on all trials. On conditioning trials the solenoids were activated for .2 sec. immediately following the cessation of the CS. For Groups XA and XB this resulted in a gradual upward movement of the right hand. The hand remained in this position for about 6 sec. and was then lowered. For Groups CA and CB the movable section was tied down and these groups consequently received only the noise produced by solenoid activation as the UCS.

Measurement of records.—The largest of the five integrated action potentials occurring during the .5-sec. interval immediately preceding the CS on each trial was used as the prestimulus level of response for that trial. The highest peak in each of the four .5-sec. intervals occurring during the CS was measured on all trials. Since solenoid activation produced a recording artifact, the unconditioned response (UCR) was determined by measuring the highest peak occurring in each of the two .5-sec. intervals following the break of the solenoids. On test trials the CR was determined by measuring the highest peak in each of the first three .5-sec. intervals following the termination of the CS. In all cases, responses during a given trial are

expressed as increments from the prestimulus level on that trial.

Results

In the interest of brevity, results will be given only in terms of changes in response within pooled trials. Figure 1 shows increments in response of extensor and flexor muscles for Groups XA and CA, while Fig. 2 shows corresponding increments for Groups XB and CB. Measures during the CS interval represent a pooling of CS interval responses on Trials 3 through 23. CR interval measures represent a pooling of CR interval responses on all test trials except Trial 1. UCR interval measures represent UCR interval measures on all conditioning trials.

From Fig. 1 and 2 it can be seen that for Groups XA and XB the level of tension of the extensor and flexor muscles steadily increased during the presentation of the CS. A definite UCR was produced in both muscles, which decayed rapidly. On test trials there was a slight increase in tension

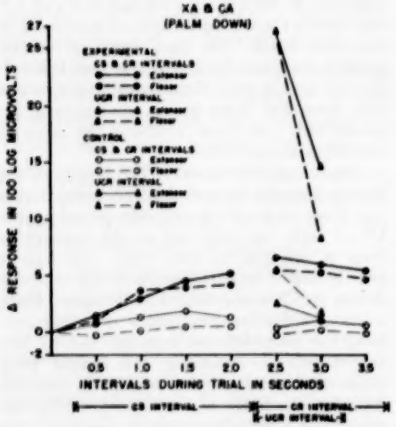

Fig. 1. Response increments within trials for Groups XA and CA. Each point represents a mean value for all *Ss* in a given group.

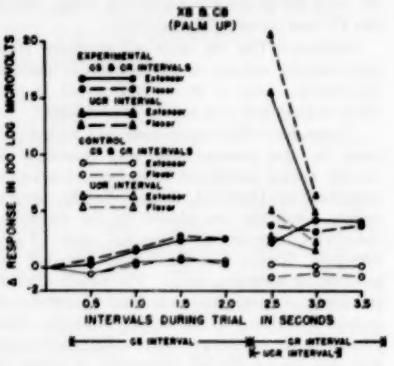

Fig. 2. Response increments within trials for Groups XB and CB. Each point represents a mean value for all *Ss* in a given group.

of both muscles during the CR interval, as compared with the CS interval. For both experimental conditions (extension movement with the palm down and flexion movement with the palm up) there was no consistent difference between extensor and flexor levels of activity.

The level of tension for both muscles in Groups CA and CB remained at about the prestimulus level during the CS and CR intervals. During the UCR interval, however, there was a slight increase in tension, indicating a response to the sound produced by solenoid activation.

The above results show graphical evidence of greater increments of tension for the experimental groups as compared to their respective control groups during all trial intervals. For a statistical test of these differences the data represented by the mean values shown in Fig. 1 and 2 were used, except that all measures within a given interval were pooled into a single measure for that interval. Since the data were found not to meet the requirements for a parametric test, a nonparametric test—the Wil-

TABLE 1

MEAN DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL GROUP INCREMENTS OF TENSION EXPRESSED IN 100 LOG MICROVOLTS

Interval	Condition			
	A (Palm Down)		B (Palm Up)	
	Extensor	Flexor	Extensor	Flexor
CS	2.14	3.31*	1.31	1.60*
CR	5.39*	5.17*	3.24	4.33*
UCR	18.82**	13.75**	8.88**	9.82**

* $P = .05$.
** $P = .01$.

coxon T test (7)—was used. Table 1 shows the mean difference between experimental and control groups for each muscle in each interval. It can be seen that in each interval at least one of the two muscles of the experimental groups manifested significantly more tension than their respective control group measures.

DISCUSSION

The experimental results give a clear-cut answer to the two major questions raised. The significantly larger muscle potentials occurring for the experimental groups, as compared with their respective control groups, during passive hand movement (UCR interval) demonstrate that action potential responses do occur in muscles directly affected by passive hand movement. This indicates that "passive movements" are not truly passive, since they result in activation of muscles in the organism.

The significantly larger responses occurring for the experimental groups during the CS and CR intervals demonstrate that conditioned muscle action potential responses can be developed in situations where the UCS is a passive movement. A CR occurred following the CS on test trials, and a conditioned anticipatory response occurred during the presentation of the CS on both conditioning and test trials. The occurrence of an increment in response during these intervals does not necessarily imply that the CR duplicates the UCR. It is possible, for example, that the tension increment is part of a general increase in tension of the organism preparatory to the UCS. Recording of action potentials from other parts of the body would be necessary to determine this. It is obvious, however, that conditioning of at least the extensor and flexor responses has taken place.

An interesting consequence of the successful conditioning produced in the present experiment is the fact that conditioning has taken place in the apparent absence of overt reinforcement. This seems to provide support for contiguity learning theories as opposed to reinforcement learning theories. A possible argument against this explanation is that the passive movement produced a "startle reflex." This would constitute the UCR, and therefore any conditioning taking place would be conditioning of a startle response rather than conditioning of a passive movement. In measuring forearm extensor and flexor action potentials occurring during a startle response, however, Davis (4) found latency and decay times of potentials accompanying the startle response which were much longer than those of the UCR produced by passive movement. It therefore appears that although passive hand movement can be successfully used as a UCS, the response to passive movement does not involve reinforcement in an obvious manner.

SUMMARY

The experiment was designed as an investigation of passive movement. By recording action potentials from muscles most likely to be involved in passive hand movement, information was obtained as to (*a*) the nature of muscular responses involved in passive hand movement, and (*b*) the conditioning of muscle action potentials resulting from these movements. Action potentials were recorded from the extensor digitorium and flexor digitorium muscles of the right forearm. In the conditioning procedure a white noise was paired with

passive hand movement for a number of conditioning trials. Interspersed among these were test trials during which the white noise was presented alone. One group of 10 *Ss* received a passive extension and another group of 10 *Ss* received a passive flexion of the right hand. For each experimental condition there was a control group of 10 *Ss* which received identical treatment except that no passive movement was given.

It was found that muscle action potential responses are produced by passive movement and that conditioned muscle action potential responses occur when passive movement is used as the UCS. The conditioning appeared to take place with no obvious reinforcement of the response.

References

1. Davis, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, **39**, 329–346.
2. Davis, R. C. The design and testing of multiple amplifiers for action potential recording. *J. exp. Psychol.*, 1943, **32**, 270–274.
3. Davis, R. C. An integrator and accessory apparatus for recording action potentials. *Amer. J. Psychol.*, 1948, **61**, 100–104.
4. Davis, R. C. Experimental investigation of the effect of a muscle tension reflex upon simple instructed movements. In *Electrophysiological aspects of aircraft control.* Formal Progress Report, Contract AF 33(038)-19630, Indiana Univer., Feb. 1, 1952.
5. Fink, J. B., & Davis, R. C. Generalization of a muscle action potential response to tonal duration. *J. exp. Psychol.*, 1951, **42**, 403–408.
6. Miller, S., & Konorski, J. Sur une forme particuliere des reflexes conditionelle. *C. R. Soc. Biol., Paris*, 1928, **99**, 1155–1157.
7. Moses, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
8. Roush, R. G. Tone burst generator. *Electronics*, 1947, **20**, 92–96.
9. Woodbury, C. G. A note on "passive" conditioning. *J. gen. Psychol.*, 1942, **27**, 359–361.

(Received October 9, 1956)

Journal of Experimental Psychology
Vol. 54, No. 4, 1957

PROBLEM SOLVING IN MULTIPLE-GOAL SITUATIONS [1]

SCARVIA B. ANDERSON

Naval Research Laboratory [2]

This experiment was designed to study the relationship between number of plausible goals with which *S*s are faced and performance on logical problems. It was hypothesized that with more than one goal, competing responses oriented toward different goals would tend to interfere with one another, and as number of goals increased amount of interference would increase. Viewing problem solving as a sequential series of behaviors (not as an "all-or-none" affair), the aim of the experiment was to specify what form interference might take in this series of behaviors, where it would be most likely to occur, and whether it would be serious enough to impede solution.

METHOD

Subjects.—The *S*s were 48 Navy enlisted men; all had General Classification Test scores of 60 (1 *SD* above the population mean of 50) or above. They were similar in age and number of years of schooling.

Material.—The calculus of propositions was used as the source of problems (**5, 6**). Each problem consisted of a series of premises from which a specified conclusion was to be drawn. Problems were presented in their abstract form, and no "meanings" were ascribed to the symbols used in them. Each *S* was trained in the use of 12 transformation rules which he could apply to the premises and the expressions derived from them in an attempt to reach the conclusion.

Two sets of problems were used: 10 practice problems (varying widely in difficulty and demanding application of all 12 of the transformation rules) and four experimental problems (of approximately equivalent difficulty). Each of the latter contained three premises and required a minimum of six steps for solution. The problem shown above the horizontal line in Table 1 is one experimental problem, as it was presented to *S*s under the one-goal condition. Under the other three experimental conditions, the information in the upper right-hand sector was changed as follows:

2-goal condition	Q or R
3-goal condition	P, Q, or R
4-goal condition	P, Q, R, or S

Under the multiple-goal conditions, only one goal was actually deducible. However, *S*s assumed all goals were deducible, and only three questioned whether it was possible to obtain even one of the goals about which they were asked.

Practice sessions.—Following training in the use of the rules of the system (see **5**), each *S* participated in seven practice sessions lasting a maximum of 30 min. each. During practice Sessions 1–3, two 15-min. problems were presented; during Sessions 4–7, one 30-min. problem was used. If *S* did not solve a problem

TABLE 1

EXAMPLE PROBLEM

(1) $(R \cdot S) \vee P$ (2) $(P \vee S) \supset (Q \vee R)$ (3) $\sim Q$	R
*(4) $P \vee (R \cdot S)$	(1) 1
*(5) $(P \vee R) \cdot (P \vee S)$	(4) 7
(6) $\sim Q \vee R$	(3) 9
(7) $\sim (Q \cdot \sim R)$	(6) 5
(8) $\sim Q \vee R$	(7) 5
(9) $Q \supset R$	(8) 6
*(10) $P \vee S$	(5) 8
*(11) $Q \vee R$	(2,10) 11
(12) $Q \supset R$	(8) 6
*(13) $\sim Q \supset R$	(11) 6
*(14) R	(13,3) 11

* Steps in minimum set (called "relevant").

[1] This article is taken from a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Maryland, 1955. The author is indebted to Dr. C. N. Cofer for serving as advisor on the thesis and to Dr. T. G. Andrews and Dr. R. C. Hackman, University of Maryland, for suggestions and criticisms; to Dr. A. H. Ryan, formerly of the Naval Research Laboratory, for support which made the study possible; and to Dr. O. K. Moore, Yale University, for criticisms, especially with respect to points of logic.

[2] Now at the Educational Testing Service.

within the time allowed, *E* demonstrated a solution. The problem-solving apparatus was identical to that used during experimental sessions.

Experimental sessions.—A randomized block design was used with two replications; each *S* received each of the four problems and each of the four goal conditions, but the combinations and orders were different for each *S* within the same group of 16. Sessions lasted a maximum of 30 min. each on four successive working days. If *S* solved a problem before 30 min. had elapsed, he was allowed to leave.

The *S* sat in a semidarkened room facing a projection screen. The *E* sat on the other side of a room divider by an overhead projector; he was not visible to *S*. At the beginning of a session, *E* turned on the projector displaying the set of premises and goal or goals appropriate to that session (e.g., the part above the horizontal line in Table 1). If more than one goal was shown, *E* said, "You will notice that there are two (three, four) letters in Sector II [the upper right-hand sector]. We will consider that you have solved the problem when you get one of those." The *S* called out to *E* suggestions for steps in terms of the rule he wished to apply and the line(s) to which he wished to apply it; e.g., "Apply Rule 11 to Lines 13 and 3" (the order which would yield Line 14, Table 1). If the rule was applicable to the line(s) specified, *E* applied it by writing the result and the "justification" on the face of the overhead projector; e.g., for Line 14, Table 1, he wrote

(14) R (13,3) 11.

If the rule was not applicable to the line or lines specified, *E* sounded a buzzer, indicating an error, and, of course, no new step was generated. The *S* was allowed to keep a rule sheet (attached to an edge-lighted clip-board) before him during all sessions, in order that his task not be primarily one of memorizing rules. During runs with the last 32 *S*s, a graphic ammeter was used to record times for errors and correct steps.

Problem-solving responses.—The kinds of problem-solving responses in which *E* was interested can be illustrated with reference to Table 1. The hypothetical *S* working on the problem took all six relevant steps in the minimum set required for solution (Lines 4, 5, 10, 11, 13, and 14) and five irrelevant steps (Lines 6, 7, 8, 9, and 12); thus he took 11 correct steps. Time and errors are not shown in Table 1, but if he made one error and worked a total of 16 min., activity level would be $(11 + 1)/16$ or .75. His first relevant step was followed by another relevant step, this second relevant step was followed by an irrelevant step, his third step was followed by an irrelevant step, etc. He introduced unnecessary material once (Line 6) and used this material in five (Lines 6, 7, 8, 9, and 12) or 45% of his steps. He used all premises.

He produced two redundancies (Lines 8 and 12). Two kinds of redundancy can be distinguished logically and the distinction is perhaps meaningful psychologically: cycles and repetitions. A cycle generally involves using an expression (e.g., Line 6) as the source of another expression (e.g., Line 7) which is in turn used to obtain the first expression again (e.g., Line 8); the number of mediating steps may vary. A repetition results if exactly the same response is made twice (e.g., Line 12 is identical with Line 9 and was obtained by the same means, applying Rule 6 to Line 8).

Results [3]

Solution.—The numbers of *S*s obtaining solution under the 1-, 2-, 3-, and 4-goal conditions were, respectively, 21, 17, 15, and 10. These numbers were subjected to a type of analysis of variance suggested by Hoyt (**3**) for use in estimating the reliability of a test. The method of partitioning variation and the results are shown in Table 2. Number of solutions is an inverse function of number of goals.

Sequence.—Figure 1 shows the proportions of the 48 *S*s who obtained each of the successive relevant steps in the minimum set leading to the one deducible goal in problems. (Irrelevant steps are not taken into account in this presentation.) It will be noted that 77, 79, 77, and 71% of *S*s obtained the first relevant step with 1, 2, 3, and 4 goals, respectively, while 44, 35, 31, and 21% of *S*s obtained Step 6 under the four goal conditions. The magnitude of the slopes of the curves gradually increases as number of goals increases.

[3] Tabulations and complete analyses of raw data and verbal reports of *S*s are included in the thesis, available through the Library of Congress on microfilm (Mic 56–261), 117 pages $1.46.

TABLE 2

ANALYSES OF VARIANCE

Source	df	Solution		Redundancy		Unnecessary Material		Use of Premises		Error	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Linear regression	1	1.2760	7.84**	1.6667	11.61***	2.3010	9.85**	.3042	2.97	.0010	—
Nonlinear regression	2	.0157	—	.0417	—	.0032	—	.2688	1.58	.3615	1.96
Goal conditious	3	.4358	2.68	.5833	4.06**	.7691	3.29*	.3472	2.04	.2413	1.31
Individuals	47	.3846	2.36***	.5246	3.65***	.2685	1.15	.3121	1.84**	.3935	2.13**
Residual	141	.1627		.1436		.2336		.1699		.1846	
Total	191										

* $P < .05$.
** $P < .01$.
*** $P < .001$.

The numbers of *S*s obtaining Relevant Steps 1, 2, 3, 4, 5, and 6 were analyzed by the Hoyt method. Goal conditions began to correlate significantly ($P < .05$) with performance at Step 4, where 22, 18, 17, and 13 *S*s obtained the step with 1–4 goals, respectively. The numbers of *S*s obtaining Step 5 were the same as the numbers of *S*s obtaining Step 6 (solution).

The empirical probabilities of problem solution, knowing *S* has been given the problem, has taken relevant Step 1, has taken Step 2, has taken Step 3, etc., are shown in Fig. 2. The first set of points ($S_0 \supset S_6$), of course, corresponds to the last set of points (Step 6) in Fig. 1. The sequential probability curve for 1 goal is a smooth increasing function, successive steps providing fairly regular increments to probability of solution. With 2, 3, and 4 goals, however, there is a very apparent change of slope in the functions. Taking Steps 1 and 2 increases somewhat the probability of solution, but a far greater increase occurs once *S* has taken Step 3 and from there on. Up to Step 5, the taking of which essentially guarantees solution for all goal conditions, the sequential probabilities of solution, knowing *S*'s progress to any point, rank in inverse relationship to the number of goals with which *S* is faced.

Figure 3 presents the empirical

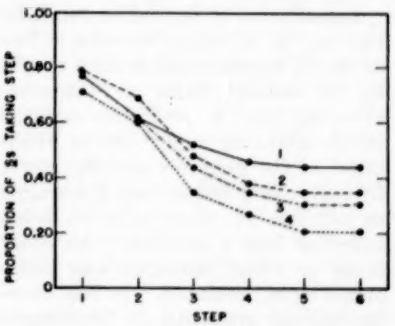

FIG. 1. Proportions of *S*s taking sequential relevant steps in 1-, 2-, 3-, and 4-goal situations.

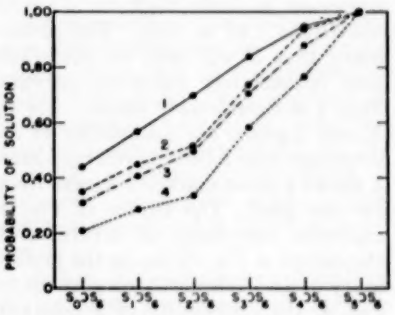

FIG. 2. Probabilities of problem solution (S_6), given the problem alone (S_0), Step 1 (S_1), etc. in 1-, 2-, 3-, and 4-goal situations.

probabilities of sequential relevant steps. For *S*s working with 2, 3, or 4 goals there is an increase in the probability of Step 2 given Step 1 over the probability of Step 1 given the problem alone, but there is a sharp decrease in the probability of Step 3 given Step 2, as compared with the preceding and succeeding probabilities of sequential steps.

The data in Fig. 4 provide some information on successive steps, *whether relevant or irrelevant.* The probability that *S*'s first step will be irrelevant is fairly high for all goal conditions (.48 to .58). The probability that *S* will take an irrelevant step immediately following relevant Step 1 is considerably lower. For 2, 3, and 4 goals, the probability of an irrelevant step following relevant Step 2 shows a more marked increase than for one goal. The curves of Fig. 4 represent essentially an inversion of the curves of Fig. 3; i.e., as the probability of an irrelevant step goes up in Fig. 4, the probability of a relevant step goes down in Fig. 3.

Curves similar in shape to those in Fig. 4 were obtained when the probabilities of relevant steps immediately following one or more irrelevant steps were plotted. The difference between the two sets of curves is in the main one of ordinate distance, the probability of a relevant step following an irrelevant step never exceeding the probability of an irrelevant step following a preceding relevant step.

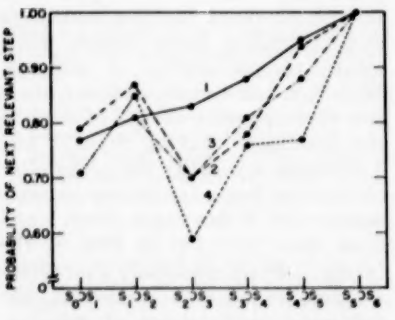

FIG. 3. Probabilities of sequential relevant steps (probability of taking Step 1 given the problem—$S_0 \supset S_1$, probability of taking Step 2 given Step 1—$S_1 \supset S_2$, etc.) in 1-, 2-, 3-, and 4-goal situations.

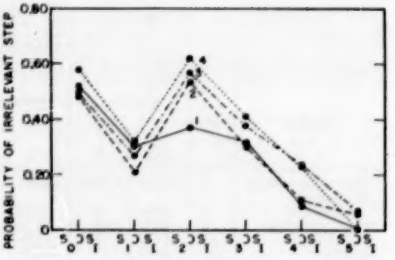

FIG. 4. Probabilities of irrelevant steps (S_I) immediately following the problem alone (S_0) or relevant steps (S_1, S_2, etc.) in 1-, 2-, 3-, and 4-goal situations.

Redundancy.—On 112 of the 192 problem protocols, no redundancies occurred. Only 13 of the 48 *S*s, however, failed to cycle or repeat at some time during their work on the four problems. The numbers of *S*s who did *not* cycle or repeat were 33, 31, 27, and 21 for 1–4 goals—a highly significant linear trend. The results of the Hoyt analysis of variance are summarized in Table 2.

Introduction of unnecessary material and use of necessary material.—Two of the 12 transformation rules taught to *S*s allowed them to introduce material into a problem, material which might or might not be duplicated in the premises initially given. In only one problem was *S* required to use one of these rules to bring anything into a problem. All other cases in which material was introduced were tabulated. In 101 problem-solving attempts, *S*s introduced unnecessary material; in 91 they did not. Thirty of the 48 *S*s under the

1-goal condition did *not* introduce unnecessary material. Numbers under the 2-, 3-, and 4-goal conditions were 25, 20, and 16. This linear trend is significant beyond the .01 level (see Table 2).

If unnecessary material was once introduced into a problem, under all goal conditions, *S*s on the average utilized it in the majority of their steps. The 18 *S*s who introduced unnecessary material in the 1-goal condition utilized it in an average of 59% of their steps. The 23 *S*s under the 2-goal condition used it on the average in 53% of their steps. The mean percentages of steps utilizing unnecessary material for the 28 *S*s under the 3-goal condition and the 32 *S*s under the 4-goal condition who introduced such material were, respectively, 57% and 63%. One, two, four, and five *S*s under the 1-, 2-, 3-, and 4-goal conditions employed unnecessary material in every one of the steps they took.

Each of the experimental problems contained three premises, the given information. Under the 1-goal condition, 35 out of 48 *S*s used all premises; the corresponding numbers under the 2-, 3-, and 4-goal conditions were, respectively, 37, 36, and 28. This variation is, however, not statistically significant (see Table 2).

Efficiency and error.—The 48 *S*s took a total of 449 correct steps under the 1-goal condition, 469 under the 2-goal condition, 461 under the 3-goal condition, and 497 under the 4-goal condition. Of these, 293, 323, 326, and 384 were irrelevant under the 1-, 2-, 3-, and 4-goal conditions, respectively. Each *S* was assigned an efficiency score, defined as the ratio of relevant steps to total correct steps, for each problem on which he worked. Many of these scores were zero, and it was advisable to use a distribution-free test of significance. Assuming, as with the Hoyt technique, that Goal × Problem and Goal × Trial interactions behaved as random variates, the median test (**4,** Formula 109) was employed. With 1 goal 28 out of the 48 scores were above the medians for their respective rows (corresponding to individuals). With 2 goals 23 were above the medians, with 3 goals 27, and with 4 goals 18. Chi square ($df = 3$) is 3.86, corresponding to a probability value between .20 and .30.

Examination of the 192 problem protocols showed that in 73 cases no errors (violations of transformation rules) were made. The numbers of *S*s *not* making any errors under the four goal conditions were 17, 18, 24, and 14, respectively. As Table 2 indicates, number of goals is not significantly related to failure to make errors.

Activity level and timing.—Numbers of responses per minute of working time for each *S* for each problem, condition, and trial were subjected to eight analyses of variance (one for each set of 48 independent scores). The over-all results indicated that activity level does not vary significantly with goal conditions, problems, trials, or interactions between goal conditions and trials or problems.

There were 20 unsolved problems each under the 1- and 2-goal conditions, 23 unsolved problems under the 3-goal condition, and 25 under the 4-goal condition for the 32 *S*s for whom detailed time records were kept. The 30-min. working time was divided into 10 3-min. intervals, and the mean cumulative percentages of responses by the end of each interval for each of the four goal conditions were computed. There is no indication that *S*s working (and not

solving) under the different conditions distributed their efforts very much differently. In all four conditions, somewhat more than half (56% to 66%) of the responses were made by the end of the half-way mark in time (15 min.).

For all *S*s who solved their problems, cumulative percentages of problems solved by the end of successive 3-min. time intervals were computed. The Kolmogorov-Smirnov (7, p. 427) test of the difference beween sample cumulative percentage polygons yielded insignificant results when the distributions for the four goal conditions were examined.

Discussion

Not all *S*s faced with one obtainable goal solved their problems. Their failures may be attributed to general and specific characteristics of the individuals and "chance." But fewer *S*s solved their problems when faced with pseudo goals in addition to the "real" goal. It is assumed that similar general and specific characteristics and "chance" affected their performances, so the differential is attributed to interference.

At what stage in the problem-solving sequence does interference occur? From the results, it is judged that it occurs after the initial restructuring steps. The *S* first rearranges one or more premises into a more workable form. The initial restructurings are probably suggested by the premises themselves and the forms of the rules and are not importantly dependent on the goal to be deduced. The majority of *S*s under all conditions get the first two relevant steps for a problem. The interval between initial restructuring steps and steps more specific to the "real" goal is the crucial interval for *S*s faced, in addition, with one or more pseudo goals. The probability of an irrelevant step occurring increases at this point, but the corresponding probability that once *S* has taken an irrelevant step he will be able to get back on the right track does not increase by so large an amount. If *S* can get by this crucial interval, if he can once again pick up the chain of steps leading to the obtainable goal, his chances of problem solution are almost as high as those for *S*s not confronted with pseudo goals.

What kinds of irrelevancies occur to distract *S*s from the path to the obtainable goal? Introduction of unnecessary material is one kind. It has been shown that once *S* has introduced unnecessary material, the probability that it will occupy a large part of his problem-solving effort is high; and the probability that unnecessary material will be introduced increases as the number of goals increases. The *S*'s behavior under these circumstances may be described as preoccupation with his own products. He tends to ignore material provided by *E*, attempting to pull a solution "out of the air." Related to *S*'s use of his own material is a tendency for the 4-goal condition toward an increase in the probability that he will not use all premises provided to him by *E*.

The increase in the probability of redundant activity as number of goals increases provides evidence for the interference hypothesis. It suggests that *S* takes a step or series of steps oriented toward one goal, is distracted by another goal, and forgets what he had done before or where his set of steps was leading. In one experiment (2) an increase in redundant activity for the experimental group did not accompany a significant decrement in number of solutions, and it was suggested that some psychological advantages of redundancy perhaps outweighed its superficial disadvantages. Another study (1), however, showed a significant though low negative correlation between number of cycles and repetitions and progress toward problem solution. In the present study, since cycles and repetitions are an increasing function of number of goals and number of solutions is a decreasing function, we cannot discount the possibility that redundancy may have served to impede solution.

Summary

Number of plausible goals with which *S* was faced was introduced as the independent variable in an experiment on human problem solving. Each of the 48 Navy *S*s worked with 1, 2, 3, and 4 goals in problems drawn from the calculus of propositions. The results, interpreted in terms of interference between responses oriented toward competing goals, indicate that (*a*) number of solutions is an inverse function of number of goals, (*b*) sequential probabilities of relevant and irrelevant problem-solving steps vary as functions of number of goals, (*c*) number of *S*s engaging in redundant activity and number of *S*s introducing unnecessary material to problems are direct functions of number of goals, (*d*) efficiency scores and number of *S*s using all given material show slight but statistically insignificant tendencies to vary as functions of number of goals, and (*e*) violations of rules for working on problems, over-all activity level, and distribution of effort over time are not related to number of goals.

References

1. Anderson, S. B. *Analysis of responses in a task drawn from the calculus of propositions.* Washington: Nav. Res. Lab., 1956. (Memo. Rep. No. 608.)
2. Anderson, S. B. *Shift in problem solving.* Washington: Nav. Res. Lab., 1955. (Memo. Rep. No. 458.)
3. Hoyt, C. J. Test reliability estimated by analysis of variance. *Psychometrika,* 1955, **10,** 64–68.
4. McNemar, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.
5. Moore, O. K., & Anderson, S. B. Modern logic and tasks for experiments on problem solving behavior. *J. Psychol.,* 1954, **38,** 151–160.
6. Moore, O. K., & Anderson, S. B. Search behavior in individual and group problem solving. *Amer. sociol. Rev.,* 1954, **19,** 702–714.
7. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

(Received October 19, 1956)

NOW AVAILABLE

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price, $1.00

Discounts for quantity orders over fifty copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

ANNOUNCEMENT

JOURNAL OF EDUCATIONAL PSYCHOLOGY

This journal will now be published by the American Psychological Association. In 1958 it will become a bimonthly; issues will appear in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscription $8.00
(Foreign, $8.50)

Single Copies, $1.50

Direct new subscriptions and renewals to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

ARE THERE GAPS IN YOUR FILES OF APA JOURNALS?

Then hear this . . .

The American Psychological Association announces a sale during the period October 1957 through March 1958. Of the following journals, all available issues in the volumes for the years preceding 1951 will be offered at a price of only 50¢ (foreign, 60¢) per issue:

American Psychologist

Journal of Abnormal & Social Psychology

Journal of Applied Psychology

Journal of Comparative & Physiological Psychology
(1947–1950 only)

Journal of Consulting Psychology

Journal of Experimental Psychology

Psychological Abstracts

Psychological Bulletin

Psychological Index
(a few complete volumes, some shopworn)

Psychological Monographs

Psychological Review

Not all issues in all volumes are available. But—ORDER NOW before more back issues go out of print. From our available stock we will complete as much of your order as possible at this reduced price and for this limited period.

Delivery:
6 to 8 weeks

No dealer or quantity discounts

After this sale, for the years preceding 1948, journals will be available only on microfilm and microcard.

Order from:

American Psychological Association
Department BB
1333 Sixteenth Street, N.W.
Washington 6, D. C.

SALE of BACK ISSUES